

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/54, 15/82, 9/10, 5/00, C12P 17/06		A2	(11) International Publication Number: WO 00/63391
			(43) International Publication Date: 26 October 2000 (26.10.00)
(21) International Application Number: PCT/US00/10368		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 14 April 2000 (14.04.00)			
(30) Priority Data: 60/129,899 15 April 1999 (15.04.99) US 60/146,461 30 July 1999 (30.07.99) US			
(71) Applicant (for all designated States except US): CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).		Published Without international search report and to be republished upon receipt of that report.	
(72) Inventors; and (75) Inventors/Applicants (for US only): SAVIDGE, Beth [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). LASSNER, Michael, W. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). WEISS, James, D. [US/US]; 800 N. Lindbergh Blvd., St. Louis, MO 63167 (US). POST-BEITENMILLER, Dusty [US/US]; 800 N. Lindbergh Blvd., St. Louis, MO 63167 (US).			
(74) Agents: SCHWEDLER, Carl, J. et al.; Calgene LLC, 1920 Fifth Street, Davis, CA 95616 (US).			
(54) Title: NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS			
(57) Abstract Nucleic acid sequences and methods are provided for producing plants and seeds having altered tocopherol content and compositions. The methods find particular use in increasing the tocopherol levels in plants, and in providing desirable tocopherol compositions in a host plant cell.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS

5

INTRODUCTION

This application claims the benefit of the filing date of the provisional Application U.S. Serial Number 60/129,899, filed April 15, 1999, and the provisional Application, U.S.
10 Serial Number 60/146,461, filed July 30, 1999.

TECHNICAL FIELD

The present invention is directed to nucleic acid and amino acid sequences and constructs, and methods related thereto.

15

BACKGROUND

Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a diverse array of greater than 22,000 isoprenoids (Connolly and Hill (1992) *Dictionary of Terpenoids*, Chapman and Hall, New York, NY). In plants, isoprenoids play essential roles in
20 particular cell functions such as production of sterols, contributing to eukaryotic membrane architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone, growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments chlorophylls and carotenoids. Although the physiological role of other plant isoprenoids is less evident, like that of the vast array of secondary metabolites, some are
25 known to play key roles mediating the adaptative responses to different environmental challenges. In spite of the remarkable diversity of structure and function, all isoprenoids originate from a single metabolic precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.* 20:525-548; and Spurgeon and Porter, (1981) in Biosynthesis of Isoprenoid Compounds, Porter and Spurgeon eds (John Wiley, New York) Vol. 1, pp1-46).

30 A number of unique and interconnected biochemical pathways derived from the isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts of higher plants. Tocopherols not only perform vital functions in plants, but are

also important from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the total quinone pool.

Tocopherols and tocotrienols (unsaturated tocopherol derivatives) are well known antioxidants, and play an important role in protecting cells from free radical damage, and in the prevention of many diseases, including cardiac disease, cancer, cataracts, retinopathy, Alzheimer's disease, and neurodegeneration, and have been shown to have beneficial effects on symptoms of arthritis, and in anti-aging. Vitamin E is used in chicken feed for improving the shelf life, appearance, flavor, and oxidative stability of meat, and to transfer tocopherols from feed to eggs. Vitamin E has been shown to be essential for normal reproduction, improves overall performance, and enhances immunocompetence in livestock animals. Vitamin E supplement in animal feed also imparts oxidative stability to milk products.

The demand for natural tocopherols as supplements has been steadily growing at a rate of 10-20% for the past three years. At present, the demand exceeds the supply for natural tocopherols, which are known to be more biopotent than racemic mixtures of synthetically produced tocopherols. Naturally occurring tocopherols are all *d*-stereoisomers, whereas synthetic α -tocopherol is a mixture of eight *d,l*- α -tocopherol isomers, only one of which (12.5%) is identical to the natural *d*- α -tocopherol. Natural *d*- α -tocopherol has the highest vitamin E activity (1.49 IU/mg) when compared to other natural tocopherols or tocotrienols. The synthetic α -tocopherol has a vitamin E activity of 1.1 IU/mg. In 1995, the worldwide market for raw refined tocopherols was \$1020 million; synthetic materials comprised 85-88% of the market, the remaining 12-15% being natural materials. The best sources of natural tocopherols and tocotrienols are vegetable oils and grain products. Currently, most of the natural Vitamin E is produced from γ -tocopherol derived from soy oil processing, which is subsequently converted to α -tocopherol by chemical modification (α -tocopherol exhibits the greatest biological activity).

Methods of enhancing the levels of tocopherols and tocotrienols in plants, especially levels of the more desirable compounds that can be used directly, without chemical modification, would be useful to the art as such molecules exhibit better functionality and bioavailability.

In addition, methods for the increased production of other isoprenoid derived compounds in a host plant cell is desirable. Furthermore, methods for the production of particular isoprenoid compounds in a host plant cell is also needed.

SUMMARY OF THE INVENTION

5 The present invention is directed to prenyltransferase (PT), and in particular to PT polynucleotides and polypeptides. The polynucleotides and polypeptides of the present invention include those derived from prokaryotic and eukaryotic sources.

 Thus, one aspect of the present invention relates to isolated polynucleotide sequences encoding prenyltransferase proteins. In particular, isolated nucleic acid sequences encoding
10 PT proteins from bacterial and plant sources are provided.

 Another aspect of the present invention relates to oligonucleotides which include partial or complete PT encoding sequences.

 It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of
15 prenyltransferase. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

 In another aspect of the present invention, methods are provided for production of prenyltransferase in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and
20 translation of prenyltransferase. The recombinant cells which contain prenyltransferase are also part of the present invention.

 In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the tocopherol content of host cells, particularly in host
25 plant cells. Plant cells having such a modified tocopherol content are also contemplated herein.

 The modified plants, seeds and oils obtained by the expression of the prenyltransferases are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

30

 Figure 1 provides an amino acid sequence alignment between ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 are performed using ClustalW.

Figure 2 provides a schematic picture of the expression construct pCGN10800.

Figure 3 provides a schematic picture of the expression construct pCGN10801.

Figure 4 provides a schematic picture of the expression construct pCGN10803.

Figure 5 provides a schematic picture of the expression construct pCGN10806.

5 Figure 6 provides a schematic picture of the expression construct pCGN10807.

Figure 7 provides a schematic picture of the expression construct pCGN10808.

Figure 8 provides a schematic picture of the expression construct pCGN10809.

Figure 9 provides a schematic picture of the expression construct pCGN10810.

Figure 10 provides a schematic picture of the expression construct pCGN10811.

10 Figure 11 provides a schematic picture of the expression construct pCGN10812.

Figure 12 provides a schematic picture of the expression construct pCGN10813.

Figure 13 provides a schematic picture of the expression construct pCGN10814.

Figure 14 provides a schematic picture of the expression construct pCGN10815.

Figure 15 provides a schematic picture of the expression construct pCGN10816.

15 Figure 16 provides a schematic picture of the expression construct pCGN10817.

Figure 17 provides a schematic picture of the expression construct pCGN10819.

Figure 18 provides a schematic picture of the expression construct pCGN10824.

Figure 19 provides a schematic picture of the expression construct pCGN10825.

Figure 20 provides a schematic picture of the expression construct pCGN10826.

20 Figure 21 provides an amino acid sequence alignment using ClustalW between the *Synechocystis* sequence knockouts.

Figure 22 provides an amino acid sequence of the ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 protein sequences from *Arabidopsis* and the slr1736, slr0926, slr1899, slr0056, and the slr1518 amino acid sequences from *Synechocystis*.

25 Figure 23 provides the results of the enzymatic assay from preparations of wild type *Synechocystis* strain 6803, and *Synechocystis* slr1736 knockout.

Figure 24 provides bar graphs of HPLC data obtained from seed extracts of transgenic *Arabidopsis* containing pCGN10822, which provides of the expression of the ATPT2 sequence, in the sense orientation, from the napin promoter. Provided are graphs for alpha, gamma, and delta tocopherols, as well as total tocopherol for 22 transformed lines, as well as a nontransformed (wildtype) control.

30

Figure 25 provides a bar graph of HPLC analysis of seed extracts from *Arabidopsis* plants transformed with pCGN10803 (35S-ATPT2, in the antisense orientation), pCGN10802

(line 1625, napin ATPT2 in the sense orientation), pCGN10809 (line 1627, 35S-ATPT3 in the sense orientation), a nontransformed (wt) control, and a empty vector transformed control.

5

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for altering (for example, increasing and decreasing) the tocopherol levels and/or modulating their ratios in
10 host cells. In particular, the present invention provides polynucleotides, polypeptides, and methods of use thereof for the modulation of tocopherol content in host plant cells.

The present invention provides polynucleotide and polypeptide sequences involved in the prenylation of straight chain and aromatic compounds. Straight chain prenyl transferases as used herein comprises sequences which encode proteins involved in the prenylation of
15 straight chain compounds, including, but not limited to, geranyl geranyl pyrophosphate and farnesyl pyrophosphate. Aromatic prenyl transferases, as used herein, comprises sequences which encode proteins involved in the prenylation of aromatic compounds, including, but not limited to, menaquinone, ubiquinone, chlorophyll, and homogentisic acid. The prenyl transferase of the present invention preferably prenylates homogentisic acid.

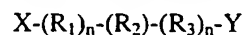
20 The biosynthesis of α -tocopherol in higher plants involves condensation of homogentisic acid and phytylpyrophosphate to form 2-methyl-6-phytylbenzoquinol that can, by cyclization and subsequent methylations (Fiedler et al., 1982, *Planta*, 155: 511-515, Soll et al., 1980, *Arch. Biochem. Biophys.* 204: 544-550, Marshall et al., 1985 *Phytochem.*, 24: 1705-1711, all of which are herein incorporated by reference in their entirety), form various
25 tocopherols. The *Arabidopsis pds2* mutant identified and characterized by Norris et al. (1995), is deficient in tocopherol and plastoquinone-9 accumulation. Further genetic and biochemical analysis suggests that the protein encoded by *PDS2* may be responsible for the prenylation of homogentisic acid. This may be a rate limiting step in tocopherol biosynthesis, and this gene has yet to be isolated. Thus, it is an aspect of the present invention to provide
30 polynucleotides and polypeptides involved in the prenylation of homogentisic acid.

Isolated Polynucleotides, Proteins, and Polypeptides

A first aspect of the present invention relates to isolated prenyltransferase polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other
 5 polynucleotide sequences closely related to such sequences and variants thereof.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding
 10 sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences
 15 that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

20 The invention also includes polynucleotides of the formula:



wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R_1 and R_3 are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence
 25 selected from the group set forth in the Sequence Listing and preferably those of SEQ ID NOs: 1, 3, 5, 7, 8, 10, 11, 13-16, 18, 23, 29, 36, and 38. In the formula, R_2 is oriented so that its 5' end residue is at the left, bound to R_1 , and its 3' end residue is at the right, bound to R_3 . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

30 The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein

5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

5 Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are
10 complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

15 Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under
20 stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate),
25 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY (1989), particularly Chapter 11.

30 The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set for in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or

a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

10 The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the prenyltransferase EST sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of prenyl transferase genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular prenyltransferase peptides, such probes may be used directly to screen gene libraries for prenyltransferase gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a prenyltransferase sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target prenyltransferase sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an prenyltransferase enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence

identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related prenyltransferase genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (See, Gould, *et al.*, *PNAS USA* (1989) 86:1934-1938.).

Another aspect of the present invention relates to prenyltransferase polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit prenyltransferase activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., *SIAM J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN)

(Coulson, *Trends in Biotechnology*, 12: 76-80 (1994); Birren, *et al.*, *Genome Analysis*, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also
 5 be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

10 Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

15 Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

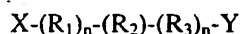
Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

20 A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:



25 wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 1000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably those encoded by the sequences provided in SEQ ID NOs: 2, 4, 6, 9, 12, 17, 19-22, 24-28, 30, 32-35, 37, and 39. In the
 30 formula, R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

5 The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide
10 of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

15 Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1
20 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

25 The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one
30 polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated

that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally
5 are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

Plant Constructs and Methods of Use

10 Of particular interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the prenyltransferase sequences of the present invention in a host plant cell. The expression constructs generally comprise a promoter functional in a host plant cell operably linked to a nucleic acid sequence encoding a prenyltransferase of the present invention and a
15 transcriptional termination region functional in a host plant cell.

A first nucleic acid sequence is "operably linked" or "operably associated" with a second nucleic acid sequence when the sequences are so arranged that the first nucleic acid sequence affects the function of the second nucleic-acid sequence. Preferably, the two sequences are part of a single contiguous nucleic acid molecule and more preferably are
20 adjacent. For example, a promoter is operably linked to a gene if the promoter regulates or mediates transcription of the gene in a cell.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid
25 operable promoters are also envisioned.

One set of plant functional promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent
30 Number 5,378, 619). In addition, it may also be preferred to bring about expression of the prenyltransferase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res.* 1:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearyl-ACP desaturase, soybean α' subunit of β -conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring prenyltransferase to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep.* 9:104-126; Clark *et al.* (1989) *J. Biol. Chem.* 264:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun.* 196:1414-1421; and, Shah *et al.* (1986) *Science* 233:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire prenyltransferase protein, or a portion thereof. For example, where antisense inhibition of a given prenyltransferase protein is desired, the entire prenyltransferase sequence is not required. Furthermore, where prenyltransferase sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a prenyltransferase encoding sequence, for example a sequence which is discovered to encode a highly conserved prenyltransferase region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726), co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the prenyltransferase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the prenyltransferase sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

The prenyltransferase constructs of the present invention can be used in transformation methods with additional constructs providing for the expression of other nucleic acid sequences encoding proteins involved in the production of tocopherols, or tocopherol precursors such as homogentisic acid and/or phytylpyrophosphate. Nucleic acid sequences encoding proteins involved in the production of homogentisic acid are known in the art, and include but not are limited to, 4-hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27) described for example, by Garcia, *et al.* ((1999) *Plant Physiol.* 119(4):1507-1516), mono or bifunctional tyrA (described for example by Xia, *et al.* (1992) *J. Gen Microbiol.* 138:1309-1316, and Hudson, *et al.* (1984) *J. Mol. Biol.* 180:1023-1051), Oxygenase, 4-hydroxyphenylpyruvate di- (9Cl), 4-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate hydroxylase; p-Hydroxyphenylpyruvate oxidase; p-Hydroxyphenylpyruvic acid hydroxylase; p-

Hydroxyphenylpyruvic hydroxylase; p-Hydroxyphenylpyruvic oxidase), 4-hydroxyphenylacetate, NAD(P)H: oxygen oxidoreductase (1-hydroxylating); 4-hydroxyphenylacetate 1-monooxygenase, and the like. In addition, constructs for the expression of nucleic acid sequences encoding proteins involved in the production of

5 phytylpyrophosphate can also be employed with the prenyltransferase constructs of the present invention. Nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate are known in the art, and include, but are not limited to geranylgeranylpyrophosphate synthase (GGPPS), geranylgeranylpyrophosphate reductase (GGH), 1-deoxyxylulose-5-phosphate synthase, 1- deoxy-D-xylulose-5-phosphate

10 reductoisomerase, 4-diphosphocytidyl-2-C-methylerythritol synthase, isopentyl pyrophosphate isomerase.

The prenyltransferase sequences of the present invention find use in the preparation of transformation constructs having a second expression cassette for the expression of additional sequences involved in tocopherol biosynthesis. Additional tocopherol biosynthesis

15 sequences of interest in the present invention include, but are not limited to gamma-tocopherol methyltransferase (Shintani, *et al.* (1998) *Science* 282(5396):2098-2100), tocopherol cyclase, and tocopherol methyltransferase.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed,

20 transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a prenyltransferase nucleic acid sequence.

Plant expression or transcription constructs having a prenyltransferase as the DNA

25 sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Particularly preferred plants for use in the methods of the present invention include, but are not limited to: *Acacia*, alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry,

30 broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom,

nectarine, nut, oat, oil palm, oil seed rape, okra, onion, orange, an ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams, and zucchini.

Most especially preferred are temperate oilseed crops. Temperate oilseed crops of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of prenyltransferase constructs in plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified content of tocopherols in plant parts having transformed plant cells.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as prenyltransferase enzymes, *in vitro* assays are performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for prenyltransferase activity. Such expression systems are known in the art and are readily available through commercial sources.

5 In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using signature sequences corresponding to conserved nucleotide and amino acid sequences of prenyltransferase can be employed to isolate equivalent, related genes from other sources such as plants and
10 microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding
15 sequences identified by any of these methods can be carried out by complementation of mutants of appropriate organisms, such as *Synechocystis*, *Shewanella*, yeast, *Pseudomonas*, *Rhodobacteria*, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

20 For the alteration of tocopherol production in a host cell, a second expression construct can be used in accordance with the present invention. For example, the prenyltransferase expression construct can be introduced into a host cell in conjunction with a second expression construct having a nucleotide sequence for a protein involved in tocopherol biosynthesis.

25 The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector
30 methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-

DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the
5 necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed,
10 where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation)
15 or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s)
20 will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) 77:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector
25 containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride, *et al.* (*Plant Mol. Biol.* (1990) 14:269-276), wherein the pRiHRI (Jouanin, *et al.*, *Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

30 Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The

particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

10 There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to
15 transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the prenyltransferase expression construct, or alternatively, transformed plants, one expressing the prenyltransferase construct and one expressing the
20 second construct, can be crossed to bring the constructs together in the same plant.

The nucleic acid sequences of the present invention can be used in constructs to provide for the expression of the sequence in a variety of host cells, both prokaryotic eukaryotic. Host cells of the present invention preferably include monocotyledenous and dicotyledenous plant cells.

25 In general, the skilled artisan is familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Maliga *et al.*,
30 *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

Methods for the expression of sequences in insect host cells are known in the art. Baculovirus expression vectors are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference). Baculovirus expression vectors are known in the art, and are described for example in Doerfler, *Curr. Top. Microbiol. Immunol.* 131:51-68 (1968); Luckow and Summers, *Bio/Technology* 6:47-55 (1988a); Miller, *Annual Review of Microbiol.* 42:177-199 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entireties of which is herein incorporated by reference)

Methods for the expression of a nucleic acid sequence of interest in a fungal host cell are known in the art. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell. Methods for the expression of DNA sequences of interest in yeast cells are generally described in "Guide to yeast genetics and molecular biology", Guthrie and Fink, eds. *Methods in enzymology*, Academic Press, Inc. Vol 194 (1991) and *Gene expression technology*, Goeddel ed, *Methods in Enzymology*, Academic Press, Inc., Vol 185 (1991).

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include, but are not limited to, viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al.*, *Nature* 273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells are well known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding epitopes into the host genome. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see*, for example, Mackett *et al*, *J Virol.* 49:857 (1984); Chakrabarti *et al.*, *Mol. Cell. Biol.* 5:3403 (1985); Moss, In: *Gene*

Transfer Vectors For Mammalian Cells (Miller and Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987); all of which are herein incorporated by reference in their entirety).

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1: Identification of Prenyltransferase Sequences

PSI-BLAST (Altschul, *et al.* (1997) *Nuc Acid Res* 25:3389-3402) profiles were generated for both the straight chain and aromatic classes of prenyltransferases. To generate the straight chain profile, a prenyl-transferase from *Porphyra purpurea* (Genbank accession 1709766) was used as a query against the NCBI non-redundant protein database. The *E. coli* enzyme involved in the formation of ubiquinone, *ubiA* (genbank accession 1790473) was used as a starting sequence to generate the aromatic prenyltransferase profile. These profiles were used to search public and proprietary DNA and protein data bases. In *Arabidopsis* seven putative prenyltransferases of the straight-chain class were identified, ATPT1, (SEQ ID NO:9), ATPT7 (SEQ ID NO:10), ATPT8 (SEQ ID NO:11), ATPT9 (SEQ ID NO:13), ATPT10 (SEQ ID NO:14), ATPT11 (SEQ ID NO:15), and ATPT12 (SEQ ID NO:16) and five were identified of the aromatic class, ATPT2 (SEQ ID NO:1), ATPT3 (SEQ ID NO:3), ATPT4 (SEQ ID NO:5), ATPT5 (SEQ ID NO:7), ATPT6 (SEQ ID NO:8). Additional prenyltransferase sequences from other plants related to the aromatic class of prenyltransferases, such as soy (SEQ ID NOs: 19-23, the deduced amino acid sequence of SEQ ID NO:23 is provided in SEQ ID NO:24) and maize (SEQ ID NOs:25-29, and 31) are also identified. The deduced amino acid sequence of ZMPT5 (SEQ ID NO:29) is provided in SEQ ID NO:30.

Searches are performed on a Silicon Graphics Unix computer using additional Bioaccelerator hardware and GenWeb software supplied by Compugen Ltd. This software and hardware enables the use of the Smith-Waterman algorithm in searching DNA and protein databases using profiles as queries. The program used to query protein databases is

profilesearch. This is a search where the query is not a single sequence but a profile based on a multiple alignment of amino acid or nucleic acid sequences. The profile is used to query a sequence data set, i.e., a sequence database. The profile contains all the pertinent information for scoring each position in a sequence, in effect replacing the "scoring matrix" used for the standard query searches. The program used to query nucleotide databases with a protein profile is tprofilesearch. Tprofilesearch searches nucleic acid databases using an amino acid profile query. As the search is running, sequences in the database are translated to amino acid sequences in six reading frames. The output file for tprofilesearch is identical to the output file for profilesearch except for an additional column that indicates the frame in which the best alignment occurred.

The Smith-Waterman algorithm, (Smith and Waterman (1981) *supra*), is used to search for similarities between one sequence from the query and a group of sequences contained in the database. E score values as well as other sequence information, such as conserved peptide sequences are used to identify related sequences.

To obtain the entire coding region corresponding to the *Arabidopsis* prenyltransferase sequences, synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends of partial cDNA clones containing prenyltransferase sequences. Primers are designed according to the respective *Arabidopsis* prenyltransferase sequences and used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) using the Marathon cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA).

Additional BLAST searches are performed using the ATPT2 sequence, a sequence in the class of aromatic prenyl transferases. Additional sequences are identified in soybean libraries that are similar to the ATPT2 sequence. The additional soybean sequence demonstrates 80% identity and 91% similarity at the amino acid sequence.

Amino acid sequence alignments between ATPT2 (SEQ ID NO:2), ATPT3 (SEQ ID NO:4), ATPT4 (SEQ ID NO:6), ATPT8 (SEQ ID NO:12), and ATPT12 (SEQ ID NO:17) are performed using ClustalW (Figure 1), and the percent identity and similarities are provided in Table 1 below.

Table 1:

ATPT2	ATPT3	ATPT4	ATPT8	ATPT12
-------	-------	-------	-------	--------

ATPT2	% Identity		12	13	11	15
	% similar		25	25	22	32
	% Gap		17	20	20	9
ATPT3	% Identity			12	6	22
	% similar			29	16	38
	% Gap			20	24	14
ATPT4	% Identity				9	14
	% similar				18	29
	% Gap				26	19
ATPT8	% Identity					7
	% similar					19
	% Gap					20
ATPT12	% Identity					
	% similar					
	% Gap					

Example 2: Preparation of Expression Constructs

- 5 A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence
- 10 CGCGATTAAATGGCGCGCCCTGCAGGCGGCCCTGCAGGGCGCGCCATTAAAT (SEQ ID NO:40) was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770, contains the pCGN7765 backbone with the napin seed specific
- 15 expression cassette from pCGN3223.

The cloning cassette, pCGN7787, essentially the same regulatory elements as pCGN7770, with the exception of the napin regulatory regions of pCGN7770 have been

replaced with the double CAMV 35S promoter and the tml polyadenylation and transcriptional termination region.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SmaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A series of turbo binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:41) and 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:42) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was excised from pCGN8618 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8622.

The plasmid pCGN8619 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:43) and 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:44) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was removed from pCGN8619 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8623.

The plasmid pCGN8620 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT -3' (SEQ ID NO:45) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:46) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8624.

The plasmid pCGN8621 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT -3' (SEQ ID NO:47) and 5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:48) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8621 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8625.

The plasmid construct pCGN8640 is a modification of pCGN8624 described above. A 938bp PstI fragment isolated from transposon Tn7 which encodes bacterial spectinomycin and streptomycin resistance (Fling et al. (1985), *Nucleic Acids Research* 13(19):7095-7106), a determinant for E. coli and Agrobacterium selection, was blunt ended with Pfu polymerase. The blunt ended fragment was ligated into pCGN8624 that had been digested with SpeI and blunt ended with Pfu polymerase. The region containing the PstI fragment was sequenced to confirm both the insert orientation and the integrity of cloning junctions.

The spectinomycin resistance marker was introduced into pCGN8622 and pCGN8623 as follows. A 7.7 Kbp AvrII-SnaBI fragment from pCGN8640 was ligated to a 10.9 Kbp

AvrII-SnaBI fragment from pCGN8623 or pCGN8622, described above. The resulting plasmids were pCGN8641 and pCGN8643, respectively.

The plasmid pCGN8644 was constructed by ligating oligonucleotides 5'-GATCACCTGCAGGAAGCTTGCGGCCGCGGATCCAATGCA-3' (SEQ ID NO:49) and
 5 5'- TTGGATCCGCGGCCGCAAGCTTCCTGCAGGT-3' (SEQ ID NO:50) into BamHI-PstI digested pCGN8640.

Synthetic oligonucleotides were designed for use in Polymerase Chain Reactions (PCR) to amplify the coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 for the preparation of expression constructs and are provided in Table 2 below.

10

Table 2:

Name	Restriction Site	Sequence	SEQ ID NO:
ATPT2	5' NotI	GGATCCGCGGCCGCGCACAATGGAGTC TCTGCTCTCTAGTTCT	51
ATPT2	3' SseI	GGATCCTGCAGGTCACTTCAAAAAA GGTAACAGCAAGT	52
ATPT3	5' NotI	GGATCCGCGGCCGCGCACAATGGCGTT TTTTGGGCTCTCCCGTGTTT	53
ATPT3	3' SseI	GGATCCTGCAGGTATTGAAAACCTT CTTCCAAGTACAAC	54
ATPT4	5' NotI	GGATCCGCGGCCGCGCACAATGTGGCG AAGATCTGTTGTT	55
ATPT4	3' SseI	GGATCCTGCAGGTCAATGGAGAGTAG AAGGAAGGAGCT	56
ATPT8	5' NotI	GGATCCGCGGCCGCGCACAATGGTACT TGCCGAGGTTCCAAAGCTTGCCCTCT	57
ATPT8	3' SseI	GGATCCTGCAGGTCACTTGTTTCTG GTGATGACTCTAT	58
ATPT12	5' NotI	GGATCCGCGGCCGCGCACAATGACTTC GATTCTCAACACT	59
ATPT12	3' SseI	GGATCCTGCAGGTCACTGTTGCGAT GCTAATGCCGT	60

The coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 were all amplified using the respective PCR primers shown in Table 2 above and cloned into the TopoTA
 15 vector (Invitrogen). Constructs containing the respective prenyltransferase sequences were digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

The sequence encoding ATPT2 prenyltransferase was cloned in the sense orientation into pCGN8640 to produce the plant transformation construct pCGN10800 (Figure 2). The ATPT2 sequence is under control of the 35S promoter.

The ATPT2 sequence was also cloned in the antisense orientation into the construct pCGN8641 to create pCGN10801 (Figure 3). This construct provides for the antisense expression of the ATPT2 sequence from the napin promoter.

The ATPT2 coding sequence was also cloned in the antisense orientation into the vector
5 pCGN8643 to create the plant transformation construct pCGN10802

The ATPT2 coding sequence was also cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10803 (Figure 4).

The ATPT4 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10806 (Figure 5). The ATPT2 coding sequence was cloned into
10 the vector pCGN864 to create the plant transformation construct pCGN10807 (Figure 6). The ATPT3 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10808 (Figure 7). The ATPT3 coding sequence was cloned in the sense orientation into the vector pCGN8640 to create the plant transformation construct pCGN10809 (Figure 8). The ATPT3 coding sequence was cloned in the antisense orientation into the vector
15 pCGN8641 to create the plant transformation construct pCGN10810 (Figure 9). The ATPT3 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10811 (Figure 10). The ATPT3 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10812 (Figure 11). The ATPT4 coding sequence was cloned into the vector pCGN8640 to create the plant transformation
20 construct pCGN10813 (Figure 12). The ATPT4 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10814 (Figure 13). The ATPT4 coding sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10815 (Figure 14). The ATPT4 coding sequence was cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10816
25 (Figure 15). The ATPT2 coding sequence was cloned into the vector pCGN???? to create the plant transformation construct pCGN10817 (Figure 16). The ATPT8 coding sequence was cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10819 (Figure 17). The ATPT12 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct pCGN10824 (Figure 18). The ATPT12 coding
30 sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10825 (Figure 19). The ATPT8 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct pCGN10826 (Figure 20).

Example 3: Plant Transformation

5 Transgenic *Brassica* plants are obtained by *Agrobacterium*-mediated transformation as described by Radke *et al.* (*Theor. Appl. Genet.* (1988) 75:685-694; *Plant Cell Reports* (1992) 11:499-505). Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) 85:5536-5540), or as described by Bent *et al.* ((1994), *Science* 265:1856-1860), or
 10 Bechtold *et al.* ((1993), *C.R.Acad.Sci, Life Sciences* 316:1194-1199). Other plant species may be similarly transformed using related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein *et al.* (*Bio/Technology* 10:286-291) may also be used to obtain nuclear transformed plants.

15

Example 4: Identification of Additional Prenyltransferases

A PSI-Blast profile generated using the *E. coli* *ubiA* (genbank accession 1790473) sequence was used to analyze the *Synechocystis* genome. This analysis identified 5 open
 20 reading frames (ORFs) in the *Synechocystis* genome that were potentially prenyltransferases; *slr0926* (annotated as *ubiA* (4-hydroxybenzoate-octaprenyl transferase, SEQ ID NO:32), *slr11899* (annotated as *ctaB* (cytochrome c oxidase folding protein, SEQ ID NO:33), *slr0056* (annotated as *g4* (chlorophyll synthase 33 kd subunit, SEQ ID NO:34), *slr1518* (annotated as *menA* (menaquinone biosynthesis protein, SEQ ID NO:35), and *slr1736* (annotated as a
 25 hypothetical protein of unknown function (SEQ ID NO:36).

To determine the functionality of these ORFs and their involvement, if any, in the biosynthesis of Tocopherols, knockouts constructs were made to disrupt the ORF identified in *Synechocystis*.

Synthetic oligos were designed to amplify regions from the 5' (5'-
 30 TAATGTGTACATTGTCGGCCTC (17365') (SEQ ID NO:61) and 5'-
 GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCACAATTCCCCGCA
 CCGTC (1736kanpr1)) (SEQ ID NO:62) and 3' (5'-AGGCTAATAAGCACAAATGGGA
 (17363') (SEQ ID NO:63) and 5'-

GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGC

GGAATTGGTTTAGGTTATCCC (1736kanpr2)) (SEQ ID NO:64) ends of the slr1736 ORF.

The 1736kanpr1 and 1736kanpr2 oligos contained 20 bp of homology to the slr1736 ORF with an additional 40 bp of sequence homology to the ends of the kanamycin resistance cassette. Separate PCR steps were completed with these oligos and the products were gel purified and combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The combined fragments were allowed to assemble without oligos under the following conditions: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min plus 5 seconds per cycle for 40 cycles using pfu polymerase in 100ul reaction volume (Zhao, H and Arnold (1997) *Nucleic Acids Res.* 25(6):1307-1308). One microliter or five microliters of this assembly reaction was then amplified using 5' and 3' oligos nested within the ends of the ORF fragment, so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21681 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The *ubiA* 5' sequence was amplified using the primers 5'-GGATCCATGGTT GCCCAAACCCCATC (SEQ ID NO:65) and 5'-GCAATGTAAACATCAGAGA TTTTGAGACACAACG TGGCTTTGGGTAAGCAACAATGACCGGC (SEQ ID NO:66). The 3' region was amplified using the synthetic oligonucleotide primers 5'-GAATTCTCAAAGCCAGCCCAGTAAC (SEQ ID NO:67) and 5'-GGTATGAGTC AGCAACACCTTCTTCACGAGGCAGACCTCAGCGGGTGCAGAAAAGGGTTTTCCC (SEQ ID NO:68). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'-CCAGTGGTTTAGGCTGTGTGGTC (SEQ ID NO:69) and 5'-CTGAGTTGGATGTATTGGATC (SEQ ID NO:70)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out.

This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21682 and used for *Synechocystis* transformation.

- Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The sl11899 5' sequence was amplified using the primers 5'-
- 5 GGATCCATGGTTACTT CGACAAAAATCC (SEQ ID NO:71) and 5'-
- GCAATGTAACATCAGAG
- ATTTTGAGACACAACGTGGCTTTGCTAGGCAACCGCTTAGTAC (SEQ ID NO:72).
- The 3' region was amplified using the synthetic oligonucleotide primers 5'-
- 10 GAATTCTTAACCCAACAGTAAAGTTCCC (SEQ ID NO:73) and 5'-
- GGTATGAGTCAGC
- AACACCTTCTTCACGAGGCAGACCTCAGCGCCGGCATTGTCTTTTACATG (SEQ ID NO:74). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector
- 15 backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'- GGAACCCTTGCAGCCGCTTC (SEQ ID NO:75)
- and 5'- GTATGCCCAACTGGTGCAGAGG (SEQ ID NO:76)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out.
- 20 This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21679 and used for *Synechocystis* transformation.

- Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The slr0056 5' sequence was amplified using the primers 5'-
- 25 GGATCCATGTCTGACACACAAAATACCG (SEQ ID NO:77) and 5'-
- GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCGCCAATACCAGCCA
- CCAACAG (SEQ ID NO:78). The 3' region was amplified using the synthetic oligonucleotide primers 5'- GAATTCTCAAAT CCCCGCATGGCCTAG (SEQ ID NO:79)
- and 5'-
- 30 GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGCCTACGGCTTG
- GACGTGTGGG (SEQ ID NO:80). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5'

and 3' oligos nested within the ends of the ORF fragment (5'-CACTTGGATTCCCCTGATCTG (SEQ ID NO:81) and 5'-GCAATACCCGCTTGGAAAACG (SEQ ID NO:82)). so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21677 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The slr1518 5' sequence was amplified using the primers 5'-GGATCCATGACCGAAT CTTCGCCCCTAGC (SEQ ID NO:83) and 5'-GCAATGTAAACATCAGAGATTTTGA GACACAACGTGGC TTTCAATCCTAGGTAGCCGAGGCG (SEQ ID NO:84). The 3' region was amplified using the synthetic oligonucleotide primers 5'-GAATTCTTAGCCCAGGCC AGCCCAGCC (SEQ ID NO:85) and 5'-GGTATGAGTCAGCAACACCTTCTTCACGA GGCAGACCTCAGCGGGAATTGATTTGTTTAATTACC (SEQ ID NO:86). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'-GCGATCGCCATTATCGCTTGG (SEQ ID NO:87) and 5'-GCAGACTGGCAATTATCAGTAACG (SEQ ID NO:88)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21680 and used for *Synechocystis* transformation.

B. Transformation of *Synechocystis*

Cells of *Synechocystis* 6803 were grown to a density of approximately 2×10^8 cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium (ATCC Medium 616) at a density of 1×10^9 cells per ml and used immediately for transformation. One-hundred microliters of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and allowed to grow for 12-18

hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing

5 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates for slr1736 and sl11899 showed complete

10 segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of these same isolates showed that the sl11899 strain had no detectable reduction in tocopherol levels. However, the strain carrying the knockout for slr1736 produced no detectable levels of tocopherol.

15 The amino acid sequences for the *Synechocystis* knockouts are compared using ClustalW, and are provided in Table 3 below. Provided are the percent identities, percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 21.

Table 3:

	Slr1736	slr0926	sl11899	slr0056	slr1518
slr1736 %identity	14	12	18	11	
%similar	29	30	34	26	
%gap	8	7	10	5	
slr0926 %identity		20	19	14	
%similar		39	32	28	
%gap		7	9	4	
sl11899 %identity			17	13	
%similar			29	29	
%gap			12	9	
slr0056 %identity				15	
%similar				31	
%gap				8	

slr1518 %identity	
%similar	
%gap	

Amino acid sequence comparisons are performed using various *Arabidopsis* prenyltransferase sequences and the *Synechocystis* sequences. The comparisons are presented in Table 4 below. Provided are the percent identities, percent similarity, and the percent gap.

5 The alignment of the sequences is provided in Figure 22.

Table 4:

	ATPT2	slr1736	ATPT3	slr0926	ATPT4	slr1899	ATPT12	slr0056	ATPT8	slr1518
ATPT2		29	9	9	8	8	12	9	7	9
		46	23	21	20	20	28	23	21	20
		27	13	28	23	29	11	24	25	24
slr1736			9	13	8	12	13	15	8	10
			19	28	19	28	26	33	21	26
			34	12	34	15	26	10	12	10
ATPT3				23	11	14	13	10	5	11
				36	26	26	26	21	14	22
				29	21	31	16	30	30	30
					12	20	17	20	11	14
slr0926					24	37	28	33	24	29
					33	12	25	10	11	9
						18	11	8	6	7
ATPT4						33	23	18	16	19
						28	19	32	32	33
							13	17	10	12
slr1899							24	30	23	26
							27	13	10	11
								52	8	11
ATPT1								66	19	26
2								18	25	23

slr0056	9	13
	23	32
	10	8
ATPT8		7
		23
		7
slr1518		

4B. Preparation of the slr1737 Knockout

The *Synechocystis* sp. 6803 slr1737 knockout was constructed by the following method. The GPS™-1 Genome Priming System (New England Biolabs) was used to insert, by a Tn7 Transposase system, a Kanamycin resistance cassette into *slr1737*. A plasmid from a *Synechocystis* genomic library clone containing 652 base pairs of the targeted orf (*Synechocystis* genome base pairs 1324051 – 1324703; the predicted orf base pairs 1323672 – 1324763, as annotated by Cyanobase) was used as target DNA. The reaction was performed according to the manufacturers protocol. The reaction mixture was then transformed into *E. coli* DH10B electrocompetant cells and plated. Colonies from this transformation were then screened for transposon insertions into the target sequence by amplifying with M13 Forward and Reverse Universal primers, yielding a product of 652 base pairs plus ~1700 base pairs, the size of the transposon kanamycin cassette, for a total fragment size of ~2300 base pairs. After this determination, it was then necessary to determine the approximate location of the insertion within the targeted orf, as 100 base pairs of orf sequence was estimated as necessary for efficient homologous recombination in *Synechocystis*. This was accomplished through amplification reactions using either of the primers to the ends of the transposon, Primer S (5' end) or N (3' end), in combination with either a M13 Forward or Reverse primer. That is, four different primer combinations were used to map each potential knockout construct: Primer S – M13 Forward, Primer S – M13 Reverse, Primer N – M13 Forward, Primer N – M13 Reverse. The construct used to transform *Synechocystis* and knockout slr1737 was determined to consist of a approximately

150 base pairs of *slr1737* sequence on the 5' side of the transposon insertion and approximately 500 base pairs on the 3' side, with the transcription of the orf and kanamycin cassette in the same direction. The nucleic acid sequence of *slr1737* is provided in SEQ ID NO:38 the deduced amino acid sequence is provided in SEQ ID NO:39.

5 Cells of *Synechocystis 6803* were grown to a density of $\sim 2 \times 10^8$ cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium at a density of 1×10^9 cells per ml and used immediately for transformation. 100 μ l of these cells were mixed with 5 μ l of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8
10 and allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES + 5 μ g/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 μ g/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10 μ g/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11
15 + kanamycin at 25 μ g/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

PCR analysis of the *Synechocystis* isolates, using primers to the ends of the *slr1737* orf, showed complete segregation of the mutant genome, meaning no copies of the wild type
20 genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of the strain carrying the knockout for *slr1737* produced no detectable levels of tocopherol.

4C. Phytol Prenyltransferase Enzyme Assays

25 [3 H] Homogentisic acid in 0.1% H_3PO_4 (specific radioactivity 40 Ci/mmol). Phytol pyrophosphate was synthesized as described by Joo, *et al.* (1973) *Can J. Biochem.* 51:1527. 2-methyl-6-phytylquinol and 2,3-dimethyl-5-phytylquinol were synthesized as described by Soll, *et al.* (1980) *Phytochemistry* 19:215. Homogentisic acid, α , β , δ , and γ -tocopherol, and tocol, were purchased commercially.

30 The wild-type strain of *Synechocystis* sp. PCC 6803 was grown in BG11 medium with bubbling air at 30°C under 50 $\mu E \cdot m^{-2} \cdot s^{-1}$ fluorescent light, and 70% relative humidity. The growth medium of *slr1736* knock-out (potential PPT) strain of this organism was

supplemented with 25 $\mu\text{g mL}^{-1}$ kanamycin. Cells were collected from 0.25 to 1 liter culture by centrifugation at 5000 g for 10 min and stored at -80°C .

Total membranes were isolated according to Zak's procedures with some modifications (Zak, *et al.* (1999) *Eur J. Biochem* 261:311). Cells were broken on a French press. Before the French press treatment, the cells were incubated for 1 hour with lysozyme (0.5%, w/v) at 30°C in a medium containing 7 mM EDTA, 5 mM NaCl and 10 mM Hepes-NaOH, pH 7.4. The spheroplasts were collected by centrifugation at 5000 g for 10 min and resuspended at 0.1 - 0.5 mg chlorophyll $\cdot\text{mL}^{-1}$ in 20 mM potassium phosphate buffer, pH 7.8. Proper amount of protease inhibitor cocktail and DNAase I from Boehringer Mannheim were added to the solution. French press treatments were performed two to three times at 100 MPa. After breakage, the cell suspension was centrifuged for 10 min at 5000 g to pellet unbroken cells, and this was followed by centrifugation at 100 000 g for 1 hour to collect total membranes. The final pellet was resuspended in a buffer containing 50 mM Tris-HCL and 4 mM MgCl_2 .

Chloroplast pellets were isolated from 250 g of spinach leaves obtained from local markets. Devined leaf sections were cut into grinding buffer (2 l /250 g leaves) containing 2 mM EDTA, 1 mM MgCl_2 , 1 mM MnCl_2 , 0.33 M sorbitol, 0.1% ascorbic acid, and 50 mM Hepes at pH 7.5. The leaves were homogenized for 3 sec three times in a 1-L blender, and filtered through 4 layers of miracloth. The supernatant was then centrifuged at 5000 g for 6 min. The chloroplast pellets were resuspended in small amount of grinding buffer (Douce, *et al* Methods in Chloroplast Molecular Biology, 239 (1982))

Chloroplasts in pellets can be broken in three ways. Chloroplast pellets were first aliquoted in 1 mg of chlorophyll per tube, centrifuged at 6000 rpm for 2 min in microcentrifuge, and grinding buffer was removed. Two hundred microliters of Triton X-100 buffer (0.1% Triton X-100, 50 mM Tris-HCl pH 7.6 and 4 mM MgCl_2) or swelling buffer (10 mM Tris pH 7.6 and 4 mM MgCl_2) was added to each tube and incubated for $\frac{1}{2}$ hour at 4°C . Then the broken chloroplast pellets were used for the assay immediately. In addition, broken chloroplasts can also be obtained by freezing in liquid nitrogen and stored at -80°C for $\frac{1}{2}$ hour, then used for the assay.

In some cases chloroplast pellets were further purified with 40%/ 80% percoll gradient to obtain intact chloroplasts. The intact chloroplasts were broken with swelling buffer, then either used for assay or further purified for envelope membranes with 20.5%/ 31.8% sucrose density gradient (Sol, *et al* (1980) *supra*). The membrane fractions were centrifuged at 100 000 g for 40 min and resuspended in 50 mM Tris-HCl pH 7.6, 4 mM MgCl_2 .

Various amounts of [^3H]HGA, 40 to 60 μM unlabelled HGA with specific activity in the range of 0.16 to 4 Ci/mmol were mixed with a proper amount of 1M Tris-NaOH pH 10 to adjust pH to 7.6. HGA was reduced for 4 min with a trace amount of solid NaBH_4 . In addition to HGA, standard incubation mixture (final vol 1 mL) contained 50 mM Tris-HCl, pH 7.6, 3-5 mM MgCl_2 , and 100 μM phytyl pyrophosphate. The reaction was initiated by addition of *Synechocystis* total membranes, spinach chloroplast pellets, spinach broken chloroplasts, or spinach envelope membranes. The enzyme reaction was carried out for 2 hour at 23°C or 30°C in the dark or light. The reaction is stopped by freezing with liquid nitrogen, and stored at -80°C or directly by extraction.

10 A constant amount of tocol was added to each assay mixture and reaction products were extracted with a 2 mL mixture of chloroform/methanol (1:2, v/v) to give a monophasic solution. NaCl solution (2 mL; 0.9%) was added with vigorous shaking. This extraction procedure was repeated three times. The organic layer containing the prenylquinones was filtered through a 20 μm filter, evaporated under N_2 and then resuspended in 100 μL of
15 ethanol.

The samples were mainly analyzed by Normal-Phase HPLC method (Isocratic 90% Hexane and 10% Methyl-t-butyl ether), and use a Zorbax silica column, 4.6 x 250 mm. The samples were also analyzed by Reversed-Phase HPLC method (Isocratic 0.1% H_3PO_4 in MeOH), and use a Vydac 201HS54 C18 column; 4.6 x 250 mm coupled with an All-tech C18
20 guard column. The amount of products were calculated based on the substrate specific radioactivity, and adjusted according to the % recovery based on the amount of internal standard.

The amount of chlorophyll was determined as described in Arnon (1949) *Plant Physiol.* 24:1. Amount of protein was determined by the Bradford method using gamma globulin as a
25 standard (Bradford, (1976) *Anal. Biochem.* 72:248)

Results of the assay demonstrate that 2-Methyl-6-Phytylplastoquinone is produced in the *Synechocystis* slr1736 knockout preparations. The results of the phytyl prenyltransferase enzyme activity assay for the slr1736 knock out are presented in Figure 23.

30 4D. Complementation of the slr1736 knockout with ATPT2

In order to determine whether ATPT2 could complement the knockout of slr1736 in *Synechocystis* 6803 a plasmid was constructed to express the ATPT2 sequence from the TAC promoter. A vector, plasmid psl1211, was obtained from the lab of Dr. Himadri Pakrasi of

Washington University, and is based on the plasmid RSF1010 which is a broad host range plasmid (Ng W.-O., Zentella R., Wang, Y., Taylor J-S. A., Pakrasi, H.B. 2000. *phrA*, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane pyrimidine dimer specific DNA photolyase. *Arch. Microbiol.* (in press)). The ATPT2 gene was isolated from the vector pCGN10817 by PCR using the following primers. ATPT2nco.pr 5'-CCATGGATTCGAGTAAAGTTGTCGC (SEQ ID NO:89); ATPT2ri.pr- 5'-GAATTCACCTTCAAAAAAGGTAACAG (SEQ ID NO:90). These primers will remove approximately 112 BP from the 5' end of the ATPT2 sequence, which is thought to be the chloroplast transit peptide. These primers will also add an NcoI site at the 5' end and an EcoRI site at the 3' end which can be used for sub-cloning into subsequent vectors. The PCR product from using these primers and pCGN10817 was ligated into pGEM T easy and the resulting vector pMON21689 was confirmed by sequencing using the m13forward and m13reverse primers. The NcoI/EcoRI fragment from pMON21689 was then ligated with the EagI/EcoRI and EagI/NcoI fragments from psl1211 resulting in pMON21690. The plasmid pMON21690 was introduced into the slr1736 *Synechocystis* 6803 KO strain via conjugation. Cells of sl906 (a helper strain) and DH10B cells containing pMON21690 were grown to log phase (O.D. 600= 0.4) and 1 ml was harvested by centrifugation. The cell pellets were washed twice with a sterile BG-11 solution and resuspended in 200 ul of BG-11. The following was mixed in a sterile eppendorf tube: 50 ul SL906, 50 ul DH10B cells containing pMON21690, and 100 ul of a fresh culture of the slr1736 *Synechocystis* 6803 KO strain (O.D. 730 = 0.2-0.4). The cell mixture was immediately transferred to a nitrocellulose filter resting on BG-11 and incubated for 24 hours at 30C and 2500 LUX(50 ue) of light. The filter was then transferred to BG-11 supplemented with 10ug/ml Gentamycin and incubated as above for ~5 days. When colonies appeared, they were picked and grown up in liquid BG-11 + Gentamycin 10 ug/ml. (Elhai, J. and Wolk, P. 1988. Conjugal transfer of DNA to Cyanobacteria. *Methods in Enzymology* 167, 747-54) The liquid cultures were then assayed for tocopherols by harvesting 1ml of culture by centrifugation, extracting with ethanol/pyrogallol, and HPLC separation. The slr1736 *Synechocystis* 6803 KO strain, did not contain any detectable tocopherols, while the slr1736 *Synechocystis* 6803 KO strain transformed with pmon21690 contained detectable alpha tocopherol. A *Synechocystis* 6803 strain transformed with psl1211(vector control) produced alpha tocopherol as well.

Example 5: Transgenic Plant Analysis

Arabidopsis plants transformed with constructs for the sense or antisense expression of the ATPT proteins were analyzed by High Pressure Liquid Chromatography (HPLC) for altered levels of total tocopherols, as well as altered levels of specific tocopherols (alpha, beta, gamma, and delta tocopherol).

Extracts of leaves and seeds were prepared for HPLC as follows. For seed extracts, 10 mg of seed was added to 1 g of microbeads (Biospec) in a sterile microfuge tube to which 500 ul 1% pyrogallol (Sigma Chem)/ethanol was added. The mixture was shaken for 3 minutes in a mini Beadbeater (Biospec) on "fast" speed. The extract was filtered through a 0.2 um filter into an autosampler tube. The filtered extracts were then used in HPLC analysis described below.

Leaf extracts were prepared by mixing 30-50 mg of leaf tissue with 1 g microbeads and freezing in liquid nitrogen until extraction. For extraction, 500 ul 1% pyrogallol in ethanol was added to the leaf/bead mixture and shaken for 1 minute on a Beadbeater (Biospec) on "fast" speed. The resulting mixture was centrifuged for 4 minutes at 14,000 rpm and filtered as described above prior to HPLC analysis.

HPLC was performed on a Zorbax silica HPLC column (4.6 mm X 250 mm) with a fluorescent detection, an excitation at 290 nm, an emission at 336 nm, and bandpass and slits. Solvent A was hexane and solvent B was methyl-t-butyl ether. The injection volume was 20 ul, the flow rate was 1.5 ml/min, the run time was 12 min (40°C) using the gradient (Table 5):

Table 5:

<u>Time</u>	<u>Solvent A</u>	<u>Solvent B</u>
0 min.	90%	10%
10 min.	90%	10%
11 min.	25%	75%
12 min.	90%	10%

Tocopherol standards in 1% pyrogallol/ ethanol were also run for comparison (alpha tocopherol, gamma tocopherol, beta tocopherol, delta tocopherol, and tocopherol (tocol) (all from Matreya).

Standard curves for alpha, beta, delta, and gamma tocopherol were calculated using Chemstation software. The absolute amount of component x is: Absolute amount of x=

Response_x x RF_x x dilution factor where Response_x is the area of peak x, RF_x is the response factor for component x (Amount_x/Response_x) and the dilution factor is 500 ul. The ng/mg tissue is found by: total ng component/mg plant tissue.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10822 for the expression of ATAT2 from the napin promoter are provided in Figure 24.

HPLC analysis results of *Arabidopsis* seed tissue expressing the ATAT2 sequence from the napin promoter (pMON10822) demonstrates an increased level of tocopherols in the seed. Total tocopherol levels are increased as much as 50 to 60% over the total tocopherol levels of non-transformed (wild-type) *Arabidopsis* plants (Figure 24).

Furthermore, increases of particular tocopherols are also increased in transgenic *Arabidopsis* plants expressing the ATAT2 nucleic acid sequence from the napin promoter. Levels of delta tocopherol in these lines are increased greater than 3 fold over the delta tocopherol levels obtained from the seeds of wild type *Arabidopsis* lines. Levels of gamma tocopherol in transgenic *Arabidopsis* lines expressing the ATAT2 nucleic acid sequence are increased as much as about 60% over the levels obtained in the seeds of non-transgenic control lines. Furthermore, levels of alpha tocopherol are increased as much as 3 fold over those obtained from non-transgenic control lines.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10803 for the expression of ATAT2 from the enhanced 35S promoter are provided in Figure 25.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

Claims

What is Claimed is:

1. An isolated nucleic acid sequence encoding a prenyltransferase.
- 5 2. An isolated nucleic acid sequence according to Claim 1, wherein said prenyltransferase is selected from the group consisting of straight chain prenyltransferase and aromatic prenyltransferase.
3. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is isolated from a eukaryotic cell source.
4. An isolated DNA sequence according to Claim 3, wherein said eukaryotic cell source is
10 selected from the group consisting of mammalian, nematode, fungal, and plant cells.
5. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from *Arabidopsis*.
6. The DNA encoding sequence of Claim 5 wherein said prenyltransferase protein is encoded by a sequence selected from the group consisting of the sequences of Figure 1.
- 15 7. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from corn.
8. The DNA encoding sequence of Claim 7 wherein said prenyltransferase protein is encoded by a sequence which includes the EST of the sequences of Figure 3.
9. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from
20 soybean.
10. The DNA encoding sequence of Claim 9 wherein said prenyltransferase protein is encoded by a sequence which includes the ESTs of the group consisting of the sequences of Figure 2 and Figure 9.
11. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is
25 isolated from a prokaryotic cell source.
12. An isolated DNA sequence according to Claim 11, wherein said prokaryotic source is *Synechocystis*.
13. A nucleic acid construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding prenyltransferase, and a
30 transcriptional termination region.
14. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding prenyltransferase is obtained from an organism selected from the group consisting of a eukaryotic organism and a prokaryotic organism.

15. A nucleic acid construct according to Claim 14, wherein said nucleic acid sequence encoding prenyltransferase is obtained from a plant source.

16. A nucleic acid construct according to Claim 15, wherein said nucleic acid sequence encoding prenyltransferase is obtained from a source selected from the group consisting of
5 *Arabidopsis*, soybean and corn.

17. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding prenyltransferase is obtained from *Synechocystis*.

18. A plant cell comprising the construct of Claim 13.

19. A method for the alteration of the tocopherol content in a host cell, comprising;
10 transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding prenyltransferase, and a transcriptional termination region.

20. The method according to Claim 19, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

15 21. The method according to Claim 20, wherein said prokaryotic cell is *Synechocystis*.

22. The method according to Claim 20, wherein said eukaryotic cell is a plant cell.

23. The method according to Claim 22, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

24. A method for producing a tocopherol compound of interest in a host cell, said method
20 comprising obtaining a transformed host cell, said host cell having and expressing in its genome: a construct having a DNA sequence encoding a prenyltransferase operably linked to a transcriptional initiation region functional in a host cell,

wherein said prenyltransferase is involved in the synthesis of tocopherols.

25. The method according to Claim 24, wherein said host cell is selected from the group
25 consisting of a prokaryotic cell and a eukaryotic cell.

26. The method according to Claim 25, wherein said prokaryotic cell is *Synechocystis*.

27. The method according to Claim 24, wherein said eukaryotic cell is a plant cell.

28. The method according to Claim 27, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

29. A method for increasing the biosynthetic flux in cell from a host cell toward
30 tocopherol production, said method comprising transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a

host cell, a DNA encoding a prenyltransferase involved in the synthesis of tocopherols, and a transcriptional termination region.

30. The method according to Claim 29, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

5 31. The method according to Claim 30, wherein said prokaryotic cell is *Synechocystis*.

32. The method according to Claim 30, wherein said eukaryotic cell is a plant cell.

33. The method according to Claim 32, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

10

ATPT2	MESS	ELVAAGG	FCWKKON	LKLHSL	ILVLRCD	SNVAKP	AR	NNLRP	DCOG	59
ATPT3	MAFFGLSRVSRRL	IKSVSP	SSALLOQOHKALSN	PVTHYTNPS	KCYPMNDNYO	MSKSGH	ELHQE	AFGVGNRYRL	ICGMS	90
ATPT4	MMWRS	VYRF	SRVSP	SLPNRPLIP	AREL	CAVNS	SEPPV	ESTAK	GITGV	69
ATPT8	MTS	ELVAAGG	FCWKKON	LKLHSL	ILVLRCD	SNVAKP	AR	NNLRP	DCOG	59
ATPT12	MTS	ELVAAGG	FCWKKON	LKLHSL	ILVLRCD	SNVAKP	AR	NNLRP	DCOG	59
ATPT2	SSLLY	ELVAAGG	FCWKKON	LKLHSL	ILVLRCD	SNVAKP	AR	NNLRP	DCOG	59
ATPT3	SVLEGR	KKDDREKSDGV	VKKASW	LDLYLPEEV	RGYAKLAR	DKP	GTWLLAV	CH	SMAL	171
ATPT4	AATATAT	TC	EISSR	UVALAGHGHY	ARCYSK	AKLSML	VVATS	GTG	MTGT	137
ATPT8	ESTDIVT	SELVR	QRGIG	NEITEM	HVALSHLL	HDDVL	DDA	TTRGVGS	LN	117
ATPT12	KVKSQT	DKAPAGG	SS	INQLLG	IKG	ASQETN	KWKIRLO	TKPV	TWP	140
ATPT2	VAAAL	KNYIVG	NO	SV	VER	CKVKNP	YIL	ELAS	GEYS	220
ATPT3	AL	URGAGC	IND	LBOD	IC	TKVDRT	KL	JAS	EL	240
ATPT4	GT	MLAASANS	NO	FF	SN	SKMKRT	ML	RE	PS	248
ATPT8	AT	AA	KNTEV	VALLAT	AN	EH	LV	TGET	ME	190
ATPT12	MSGPC	UTGY	TO	INDWY	DRD	IC	AINEP	YRE	IP	227
ATPT2	ELRWR	RFAL	AM	CI	AVRA	IF	QAFV	LHI	QTH	300
ATPT3	AMK	ETP	PO	ET	IN	MGAL	ET	AN	KES	320
ATPT4	LKQLH	PI	NT	W	GA	PP	LT	MA	AA	340
ATPT8	LAFEG	RL	GL	AF	OH	I	DD	IT	ETS	360
ATPT12	PPUK	KONG	W	GN	FA	LG	AS	YS	IS	380
ATPT2	FS	TI	BO	KR	V	TS	VL	Q	MA	400
ATPT3	VG	TS	TA	RG	ND	NK	L	HT	FG	420
ATPT4	VAL	NC	FY	MI	PT	GH	AY	DW	GL	440
ATPT8	VD	AL	EY	GL	SK	CR	Q	LA	RE	460
ATPT12	EG	LOS	LP	AF	GT	ET	AK	IC	VC	480
ATPT2	MF	WK	LF	YAE	YLL	LP	FLK	393		
ATPT3	GAI	IF	SG	VLG	RS	F	393			
ATPT4	TNS	V	SG	EV	KTOR	KRVA	PPV	AY	ASA	407
ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	ATPT8	431
ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	ATPT12	437

Figure 1

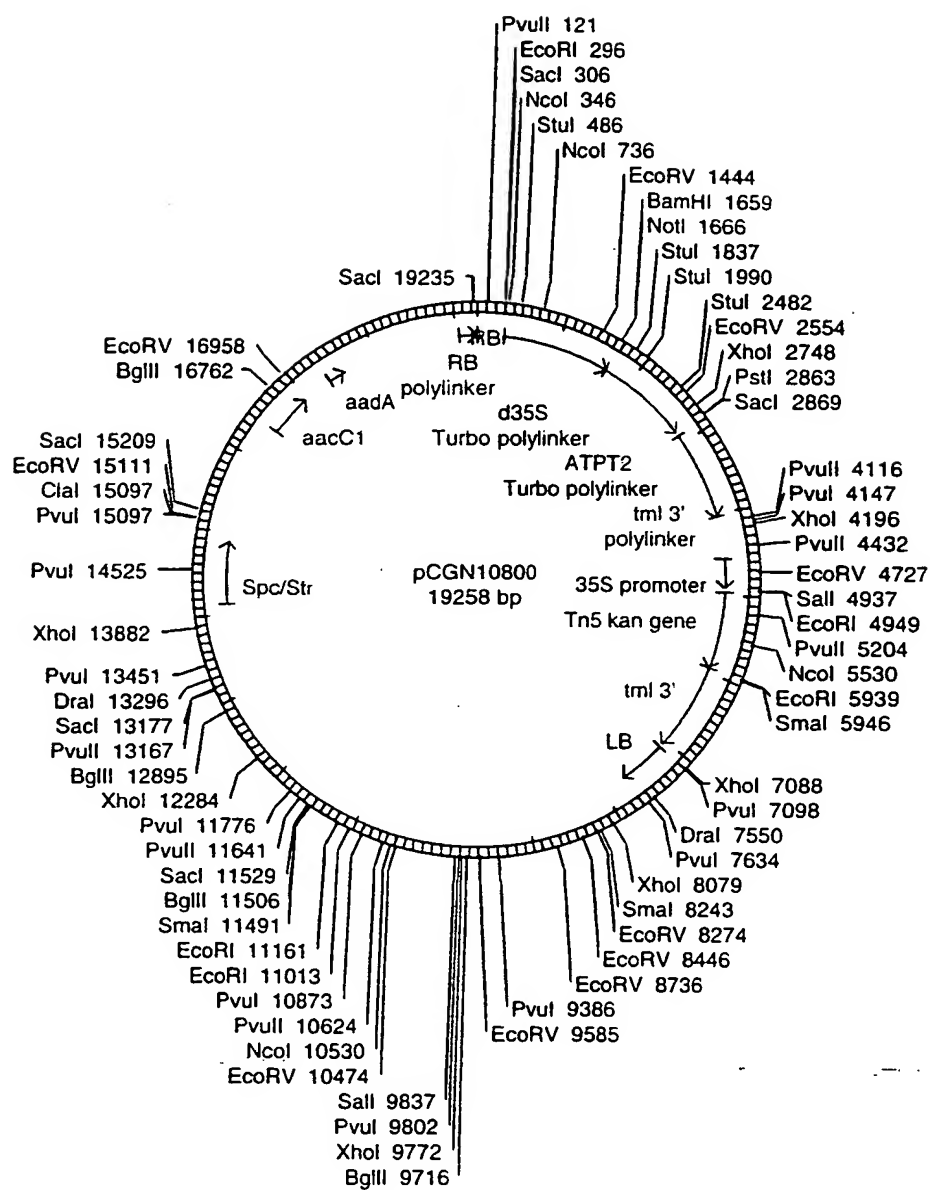


Figure 2

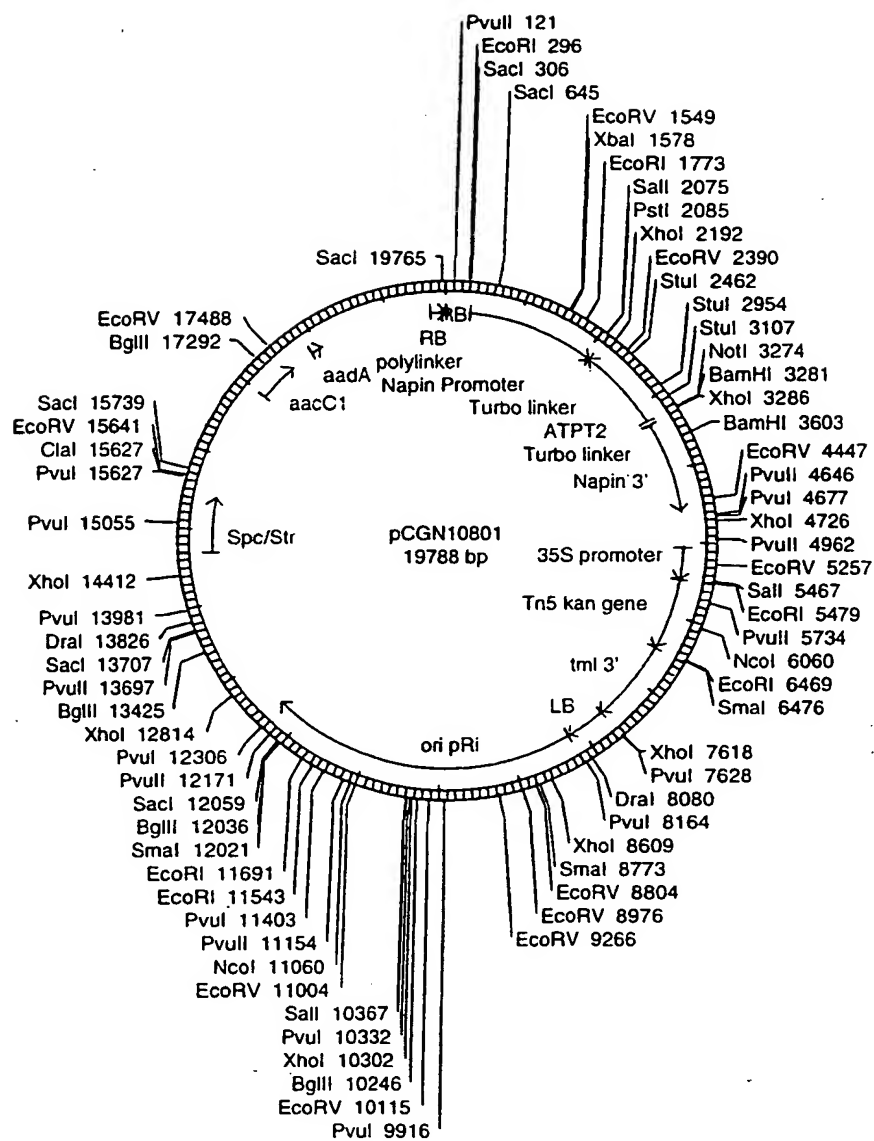


Figure 3

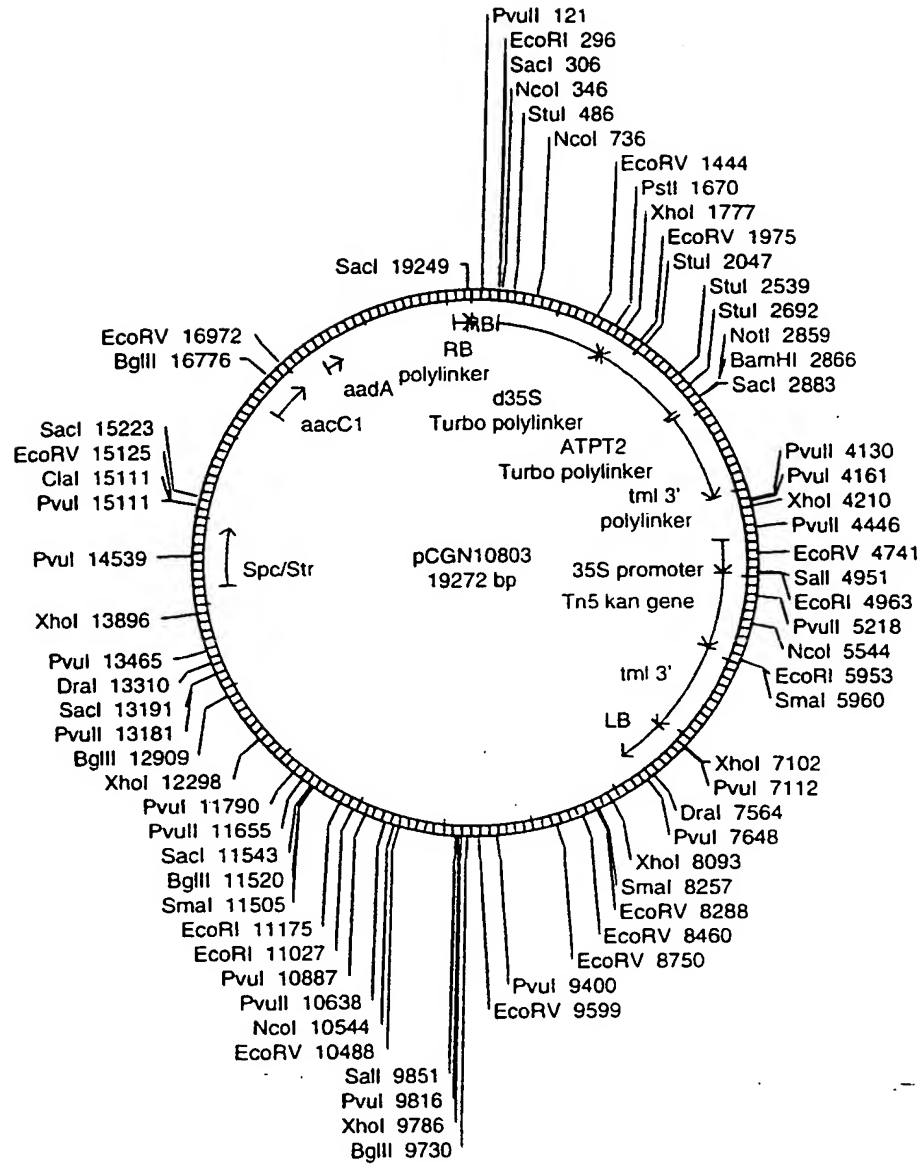


Figure 4

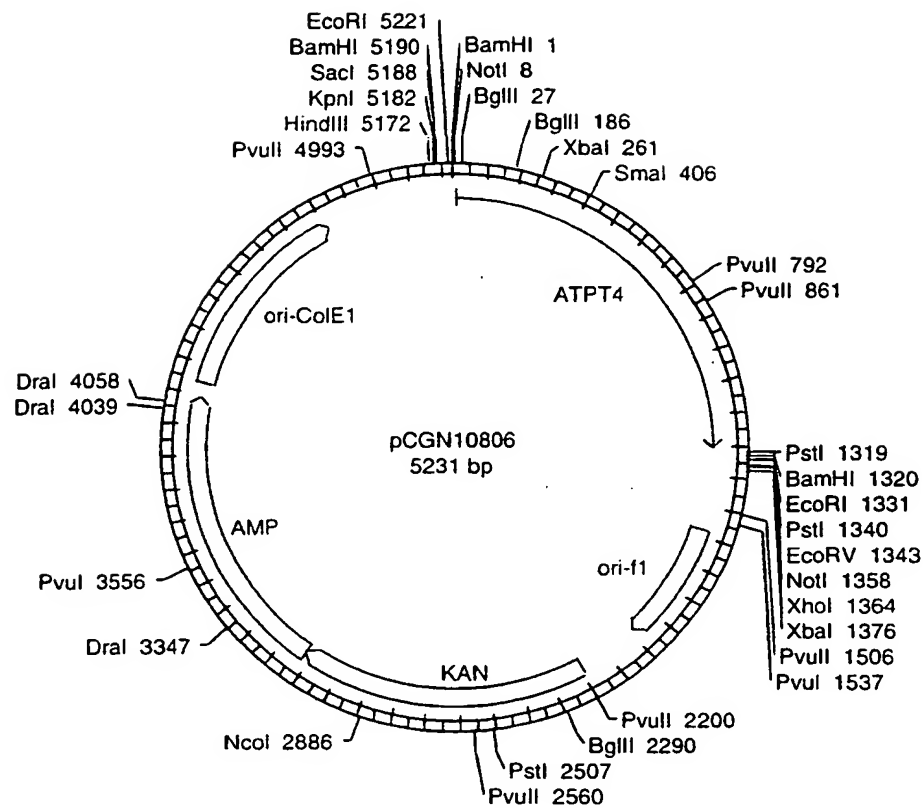


Figure 5

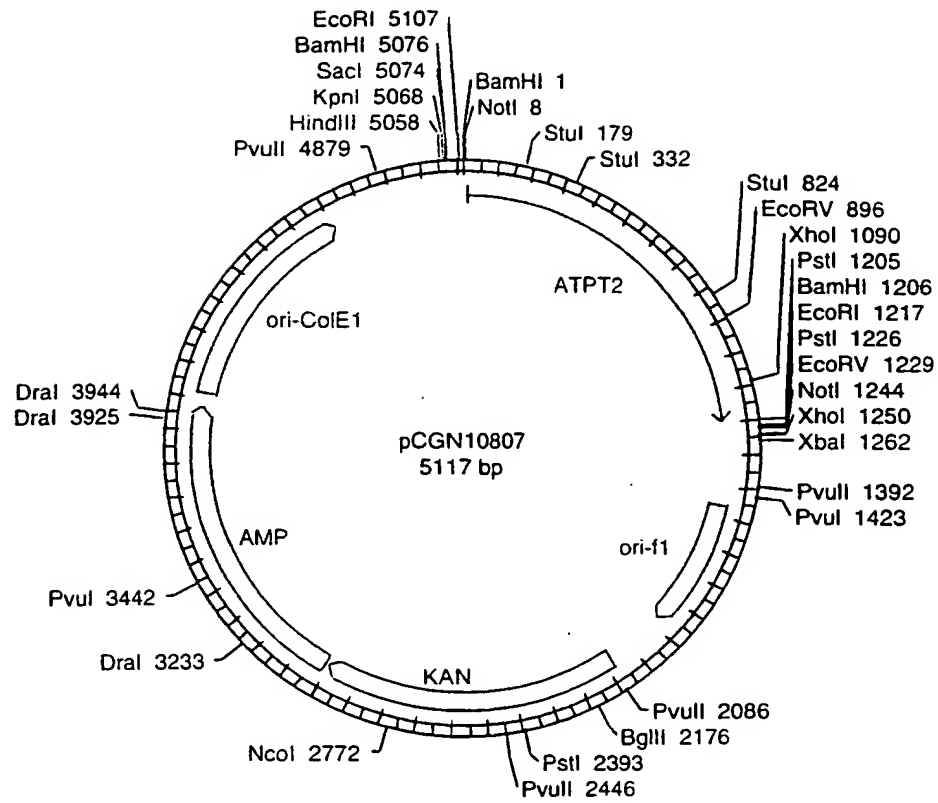


Figure 6

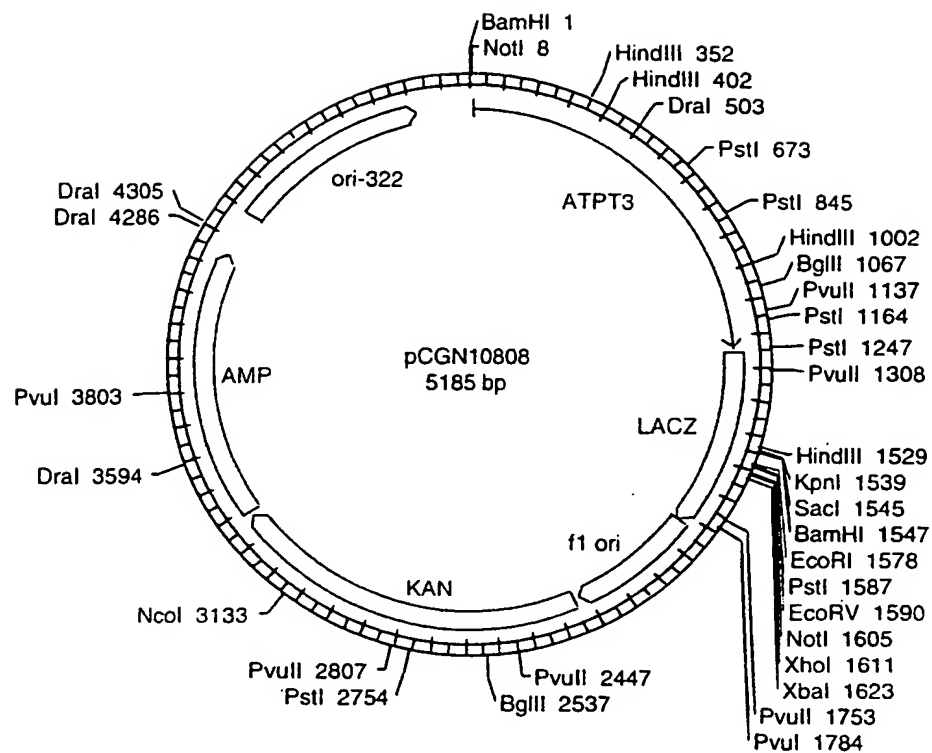


Figure 7

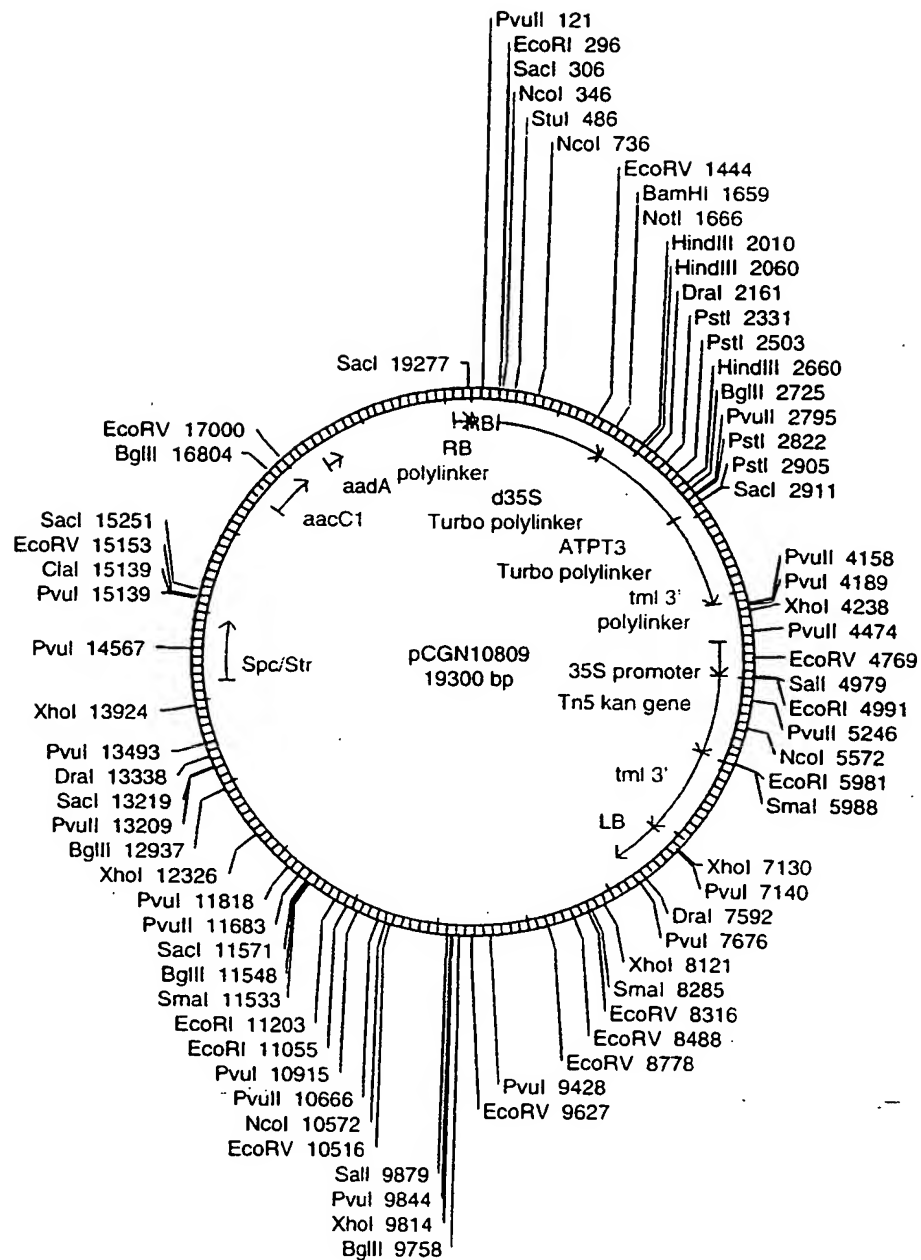


Figure 8

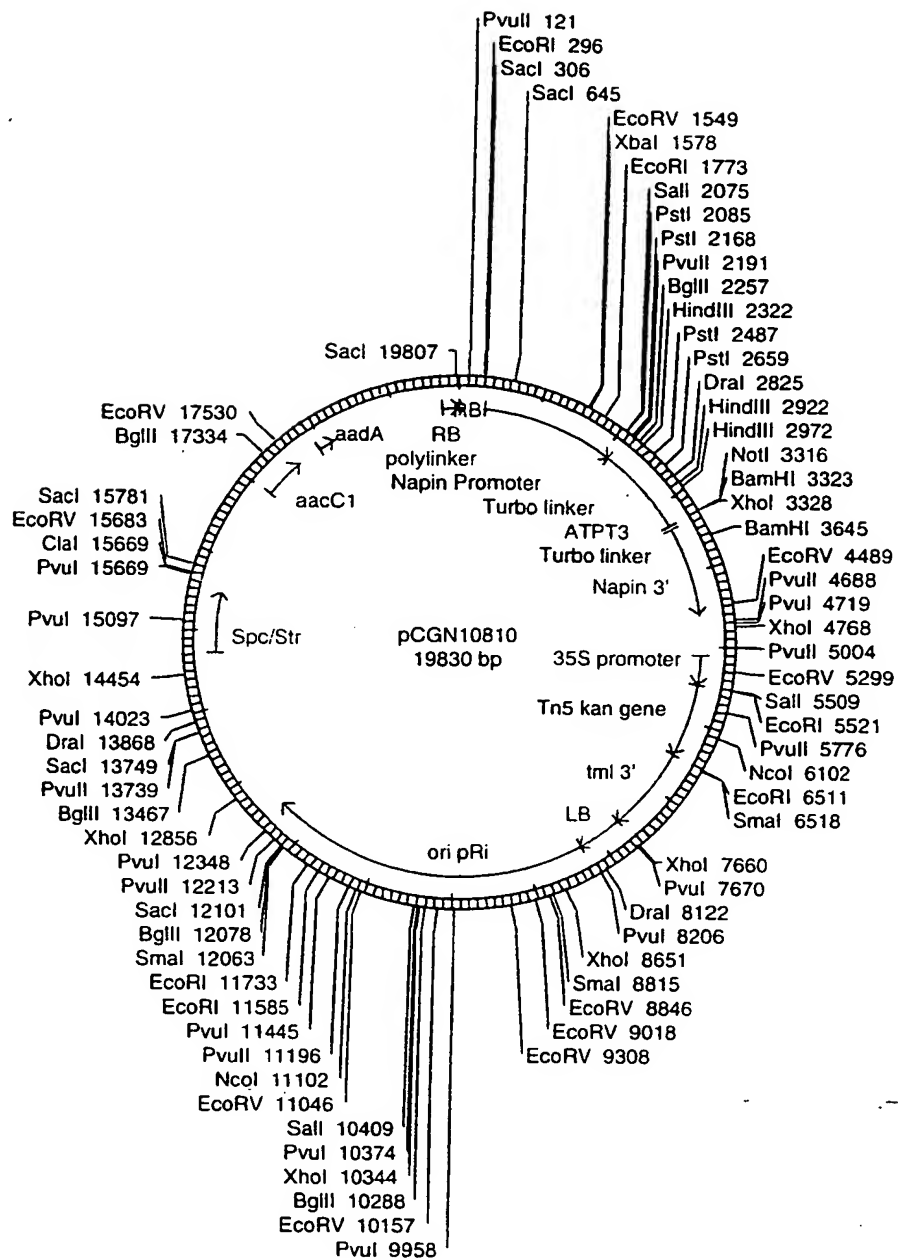


Figure 9

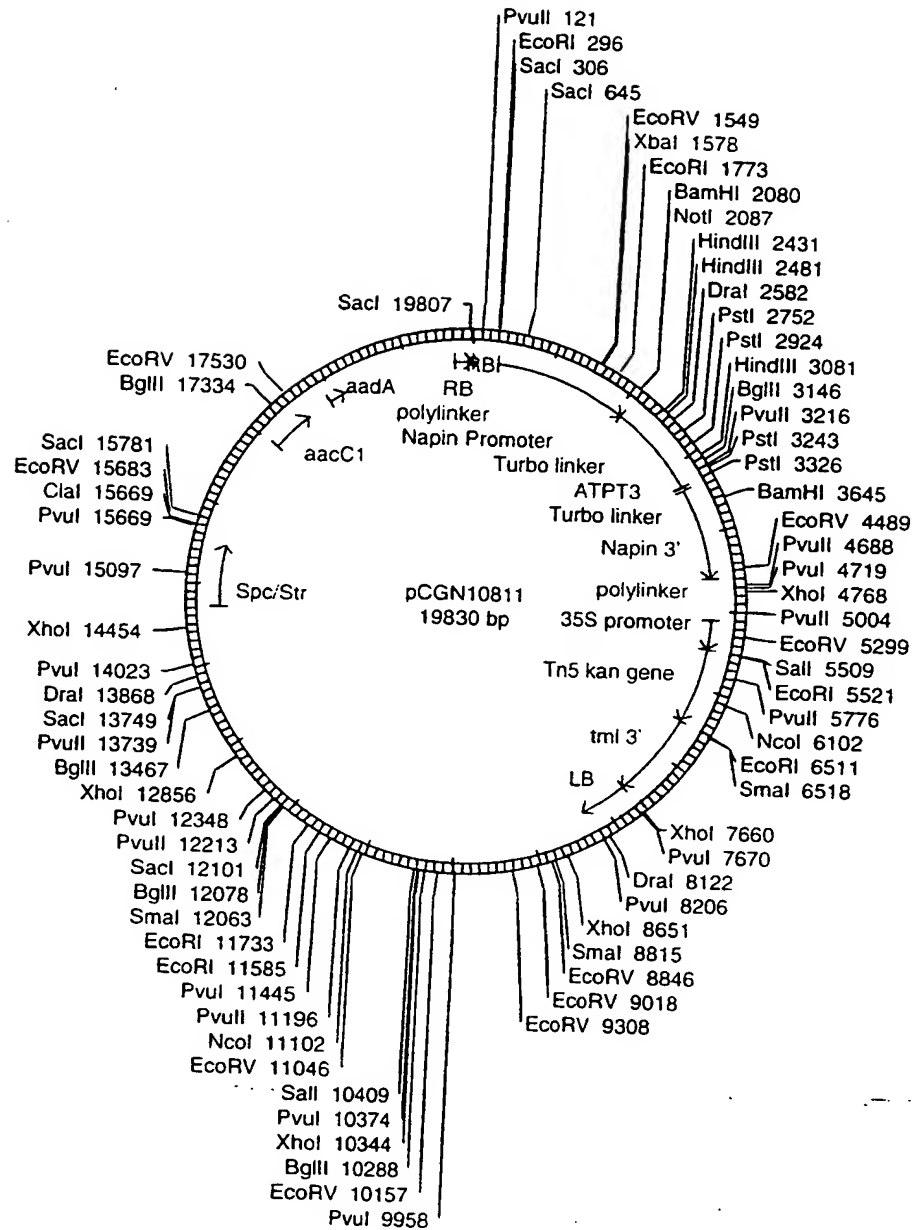


Figure 10

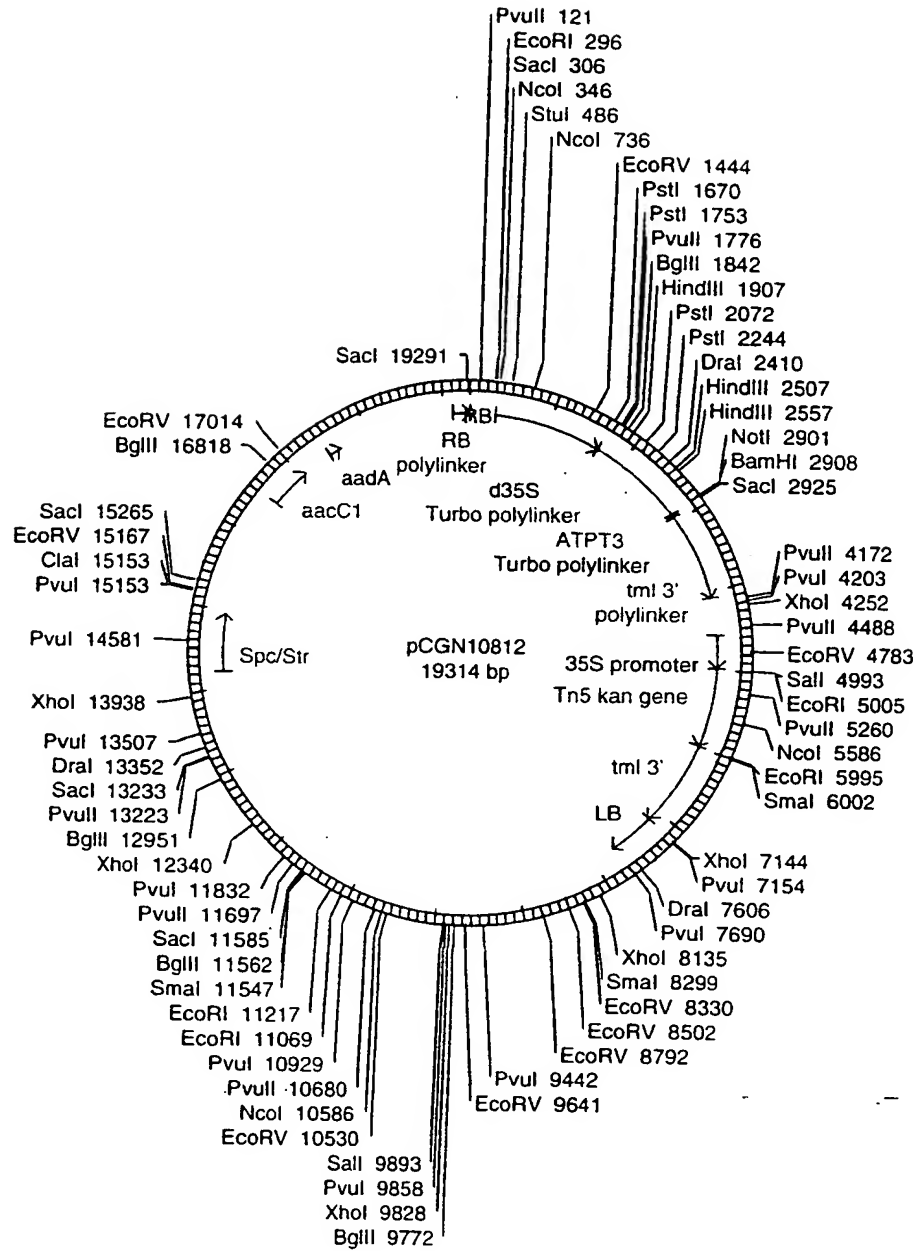


Figure 11

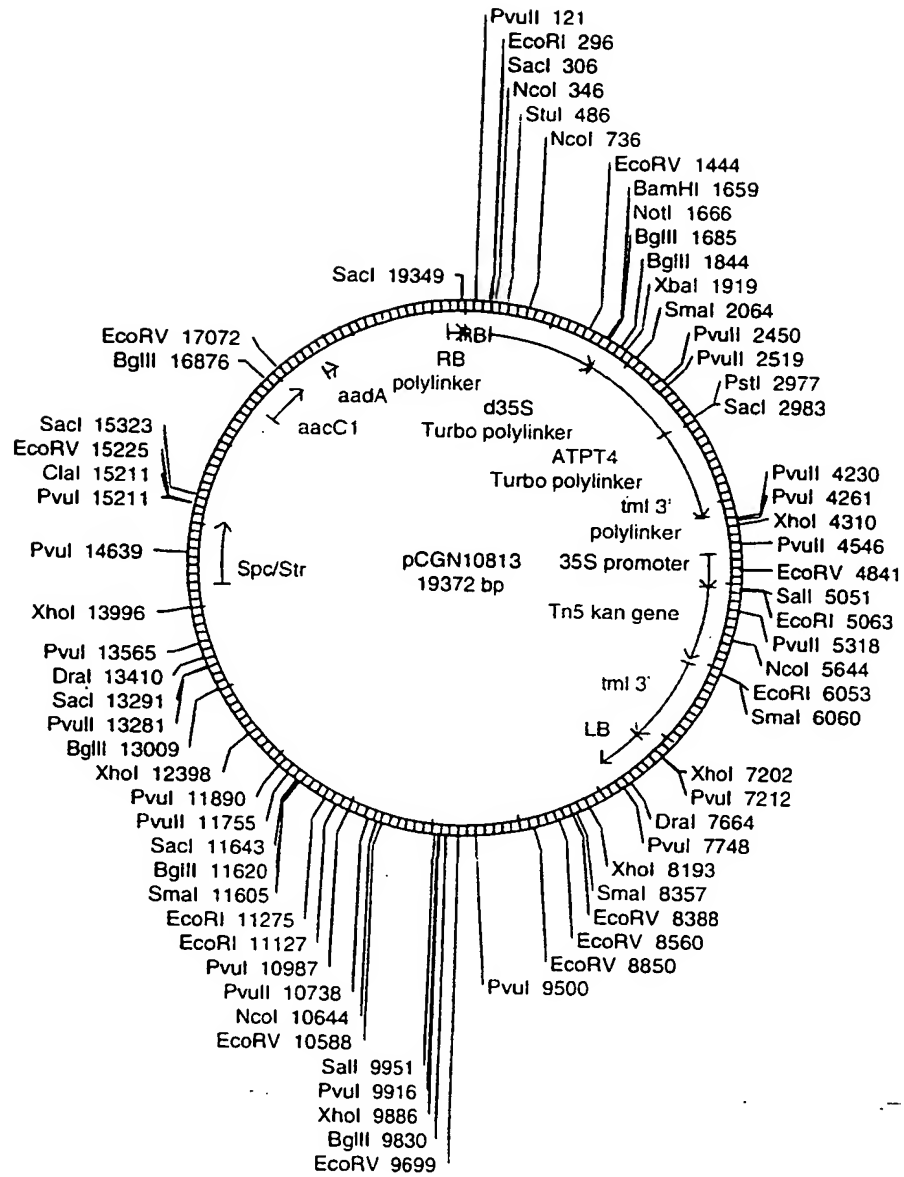


Figure 12

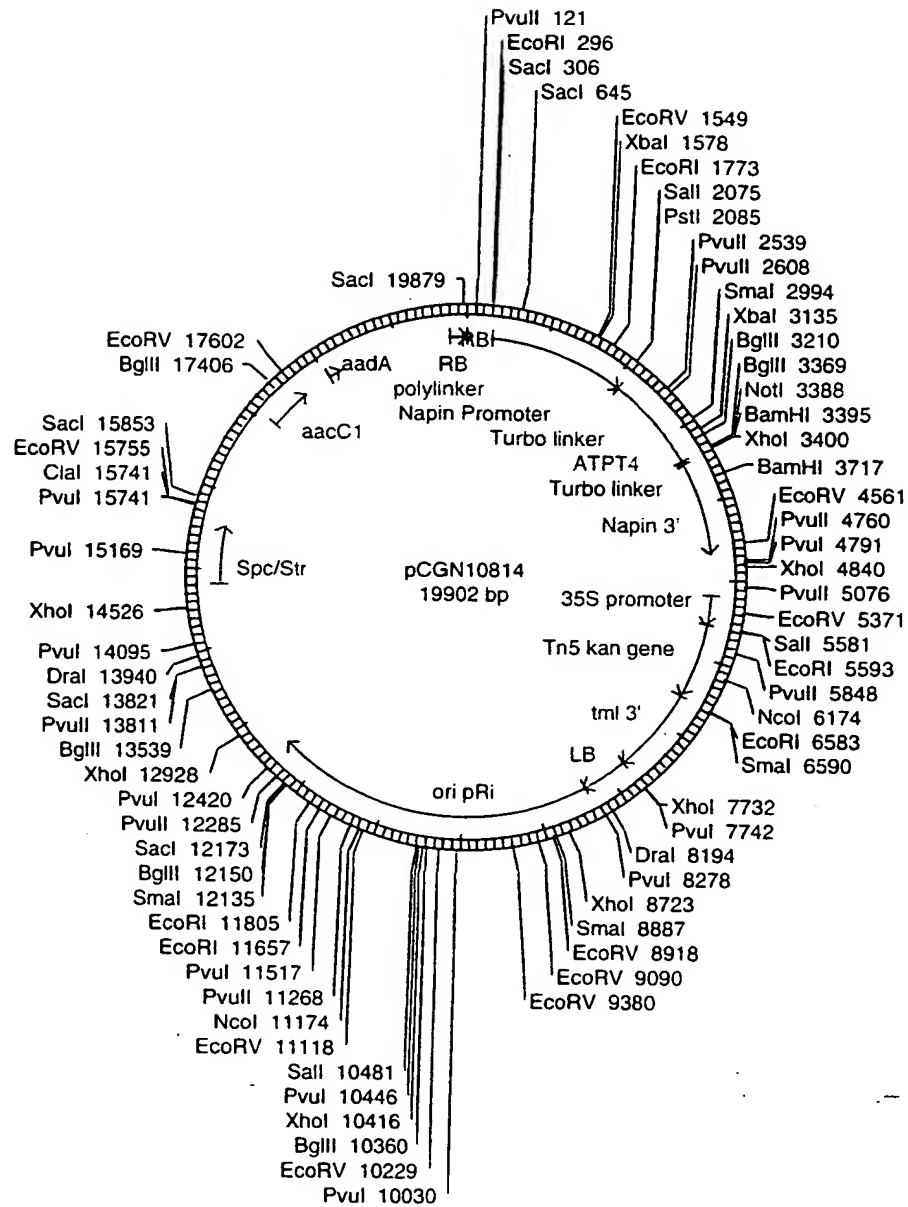


Figure 13

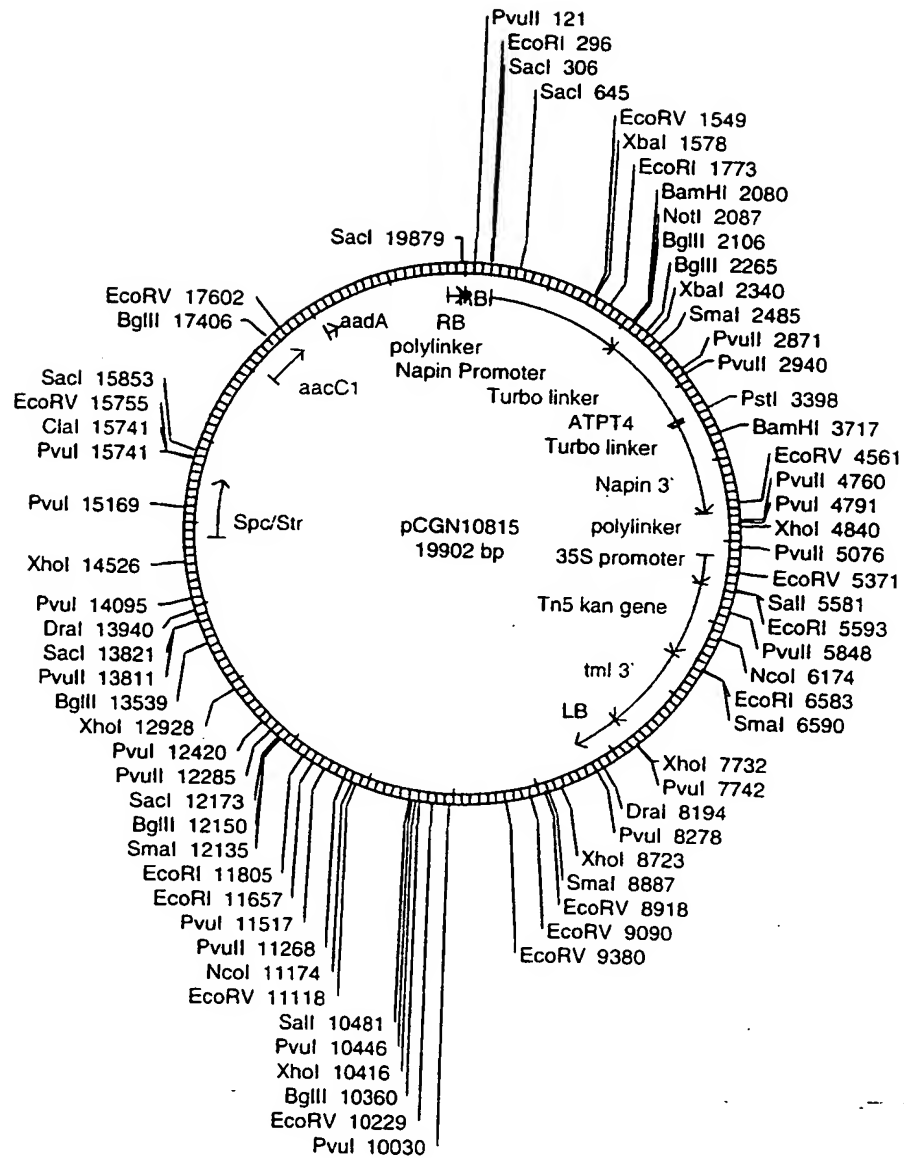


Figure 14

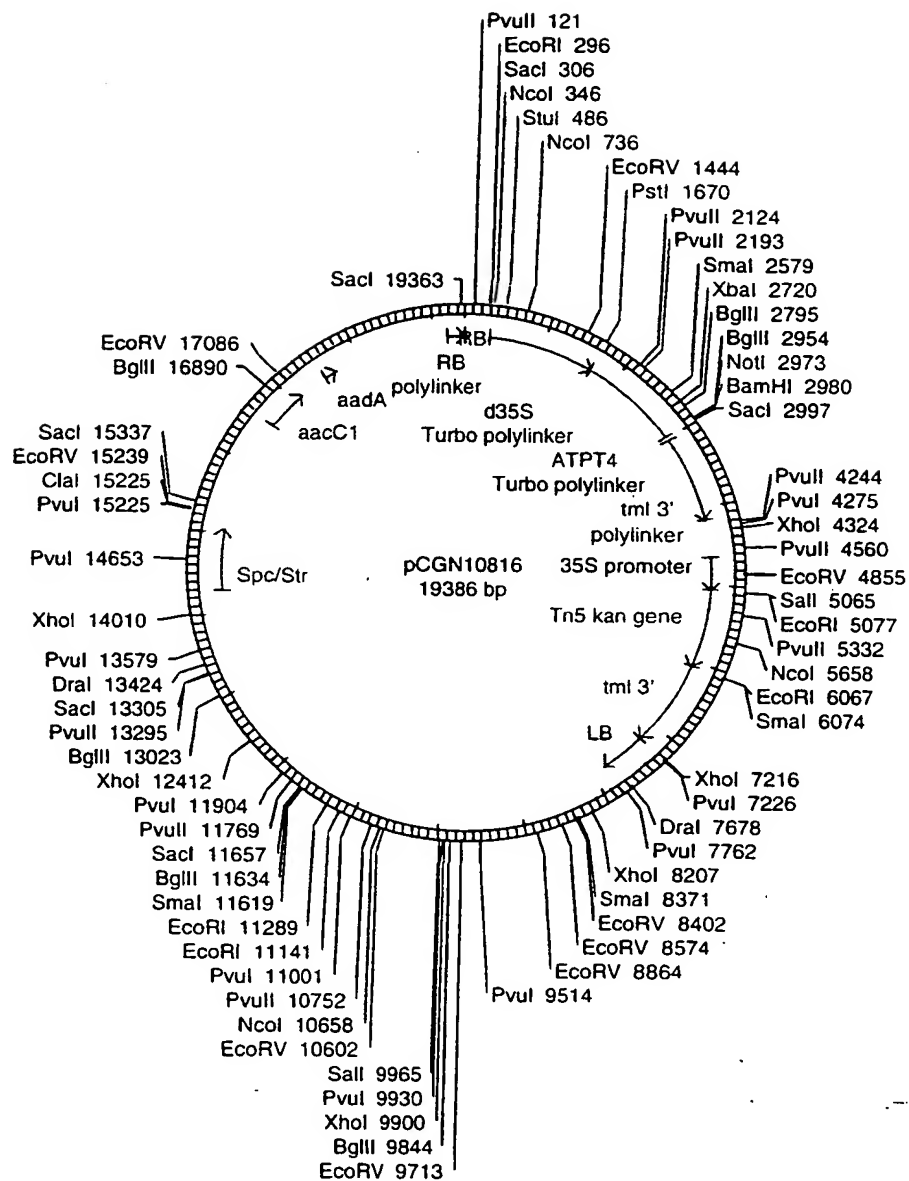


Figure 15

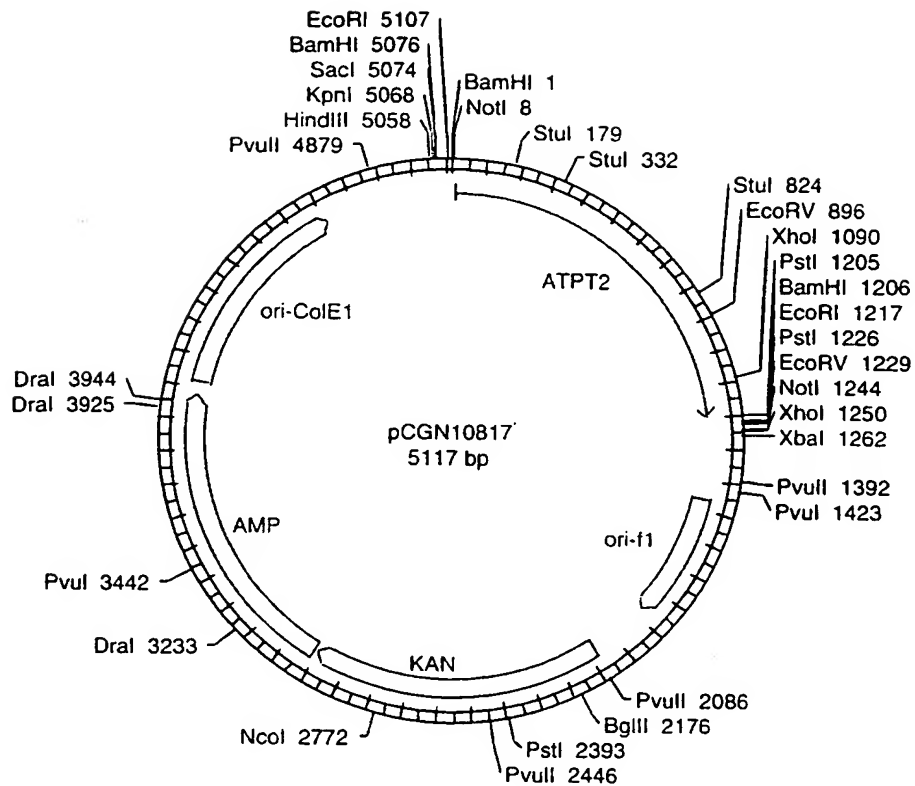


Figure 16

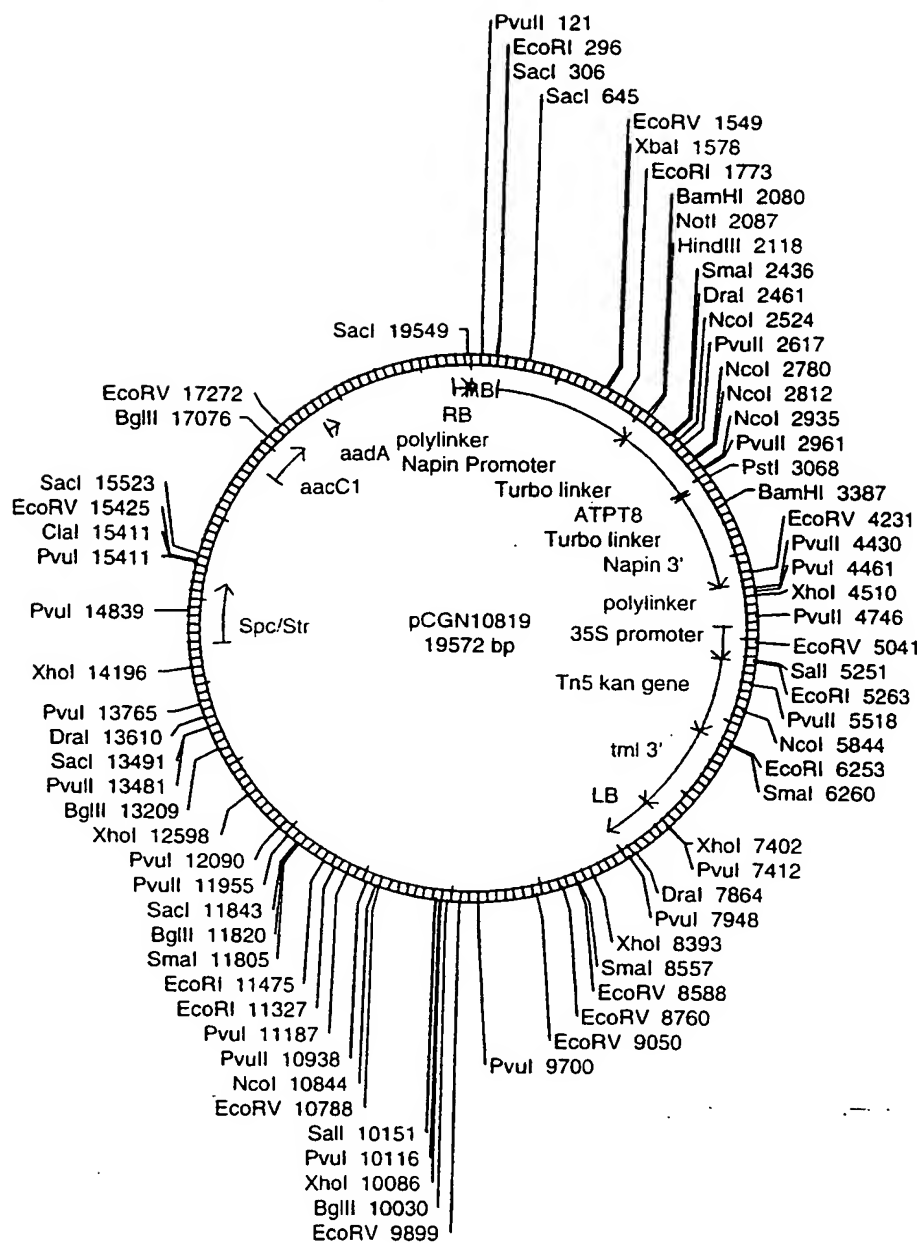
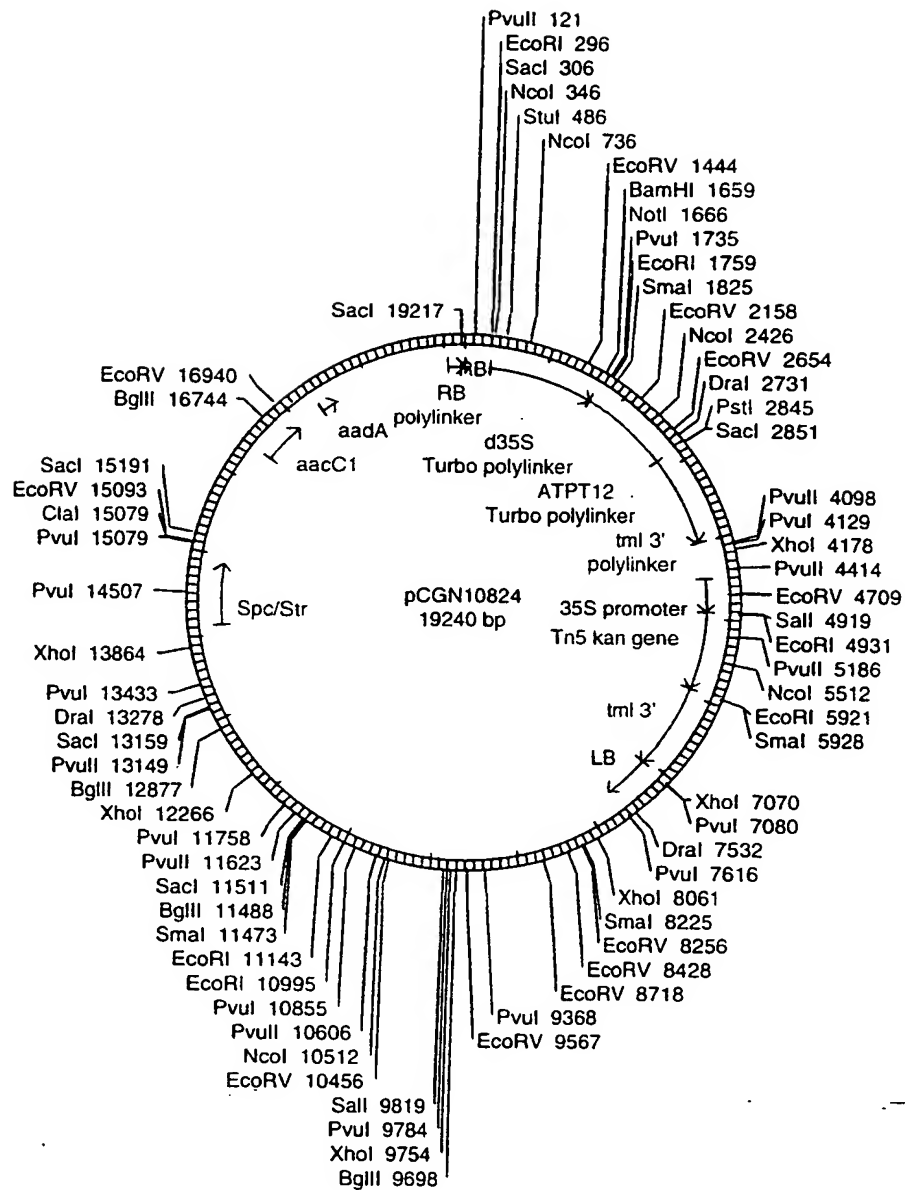


Figure 17



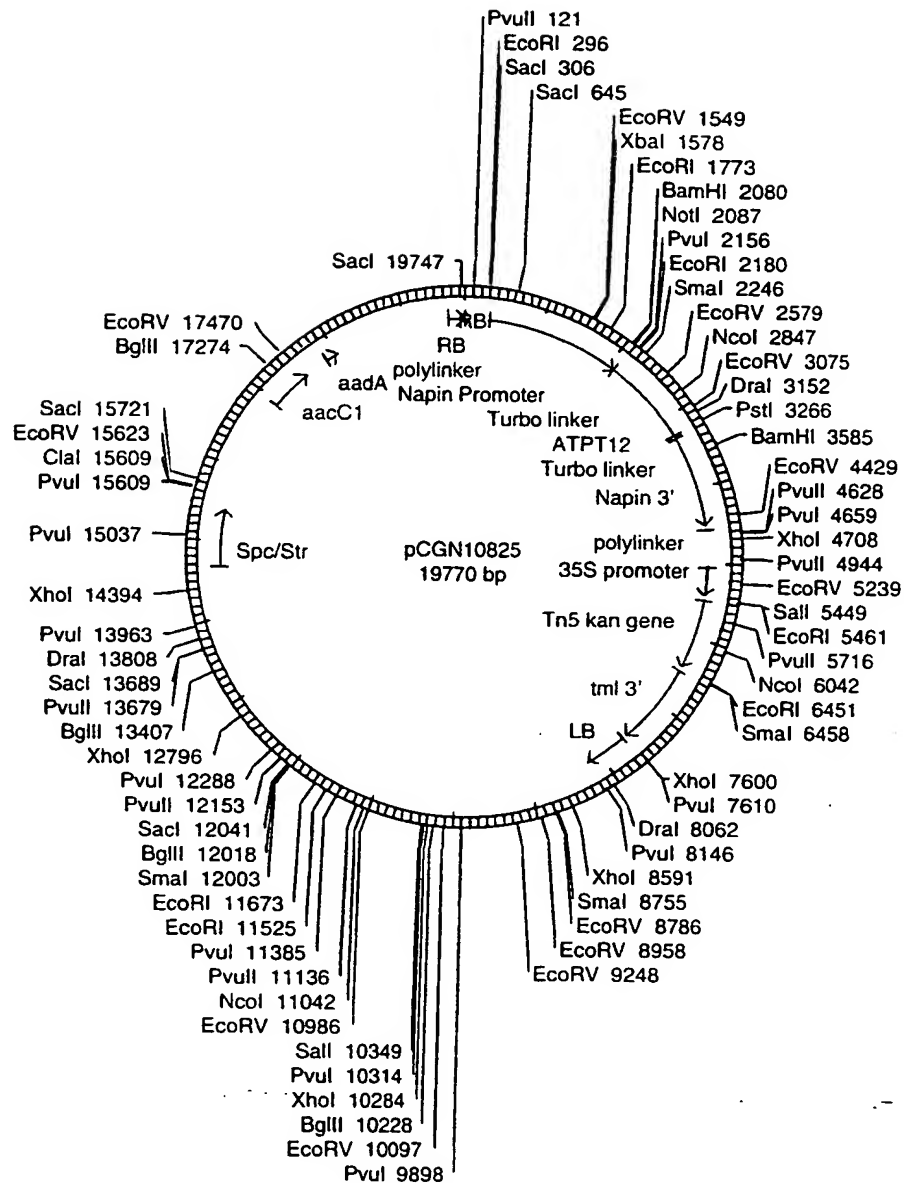


Figure 19

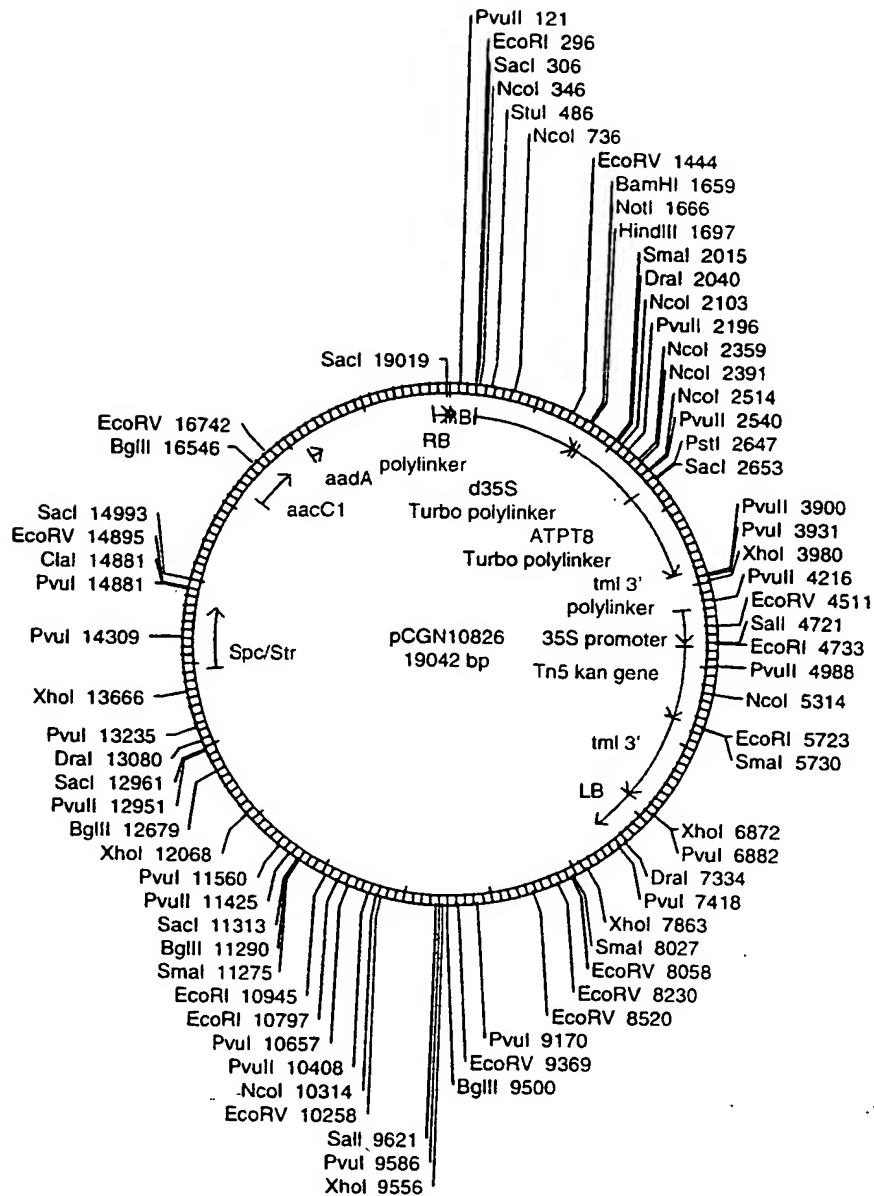


Figure 20

SLR1736 : MATIQAFWR-----FSRPHTTIGTTLS-----* 20 40 60 80 *
 SLR0926 : MVAQHPSS---PPLWLTIIYLRWHKP-----* 64
 SLR1899 : MVTSEKKIHRHDSMGVCKSYQLTKP-----* 68
 SLR0056 : MSDQNTQG--NOAMRQLLGKGAARDESSINKRIQLRKWTWIPLIINGVYCEAASSGGYIWSVEDFLKATCMISGPEMTGCTQTL : 71
 SLR1518 : MTESSPLAP--STAPATRKLMIAAIRP-----* 71
 N 5D Did 100 120 140 160 180
 SLR1736 : NQLEWDVDIRINKPNLPLANGDFSTAGRWIVGCGVAGSAIRVSGGLAGGLTVGIS--DITGTYSYVPPPRUKRFRSLLAALCETVRGH : 152
 SLR0926 : NQLEWDVDIRINKPNLPLANGDFSTAGRWIVGCGVAGSAIRVSGGLAGGLTVGIS--DITGTYSYVPPPRUKRFRSLLAALCETVRGH : 150
 SLL1899 : NCEYDQDIDYENLRTRARRIPFGKV--SPRHATIFALALVESFALDATGVNVLSCG---DASSIVFYVYTHWLRKRTAQNMAVIG-GA : 156
 SLR0056 : NCEYDQDIDYENLRTRARRIPFGKV--SPRHATIFALALVESFALDATGVNVLSCG---DASSIVFYVYTHWLRKRTAQNMAVIG-GA : 177
 SLR1518 : NQLEWDVDIRINKPNLPLANGDFSTAGRWIVGCGVAGSAIRVSGGLAGGLTVGIS--DITGTYSYVPPPRUKRFRSLLAALCETVRGH : 157
 N 5D Did 200 220 240 260
 SLR1736 : VVNLGLFEPFRIGLGYPPPTITPTWVTHFTLTPVVAIAIFKQVPMEGDQKFKQNTTIOICKANNPRGTLTTLTG-CYFAMALMGDWA : 241
 SLR0926 : AMGEAVVLSNS---AVTGDIDATWNGATVFWIIGFTVYADREDDRRRGGVSSALFFG-CYFGEAVGFFA--ETIGCLFYIGMI : 234
 SLL1899 : AGHIPPPLVGVW---AVTGDIDATWNGATVFWIIGFTVYADREDDRRRGGVSSALFFG-CYFGEAVGFFA--ETIGCLFYIGMI : 241
 SLR0056 : GAEVIALPWWAG-HALFEEAATPTMMLIIYSAGCGAVVNDKSKVEGDEQDQGLKSPVWFC-IGTAAMICAMM--DVEQAGAGLYLI : 263
 SLR1518 : LITFGPTATAAAYYSQSQSFBNLLTSPVFGVSTAIILFCSHFHQVEDDLAAGKSPIMVRLG-TXEGSQVLTLSAVSHYETITALIGVECH : 246
 N 5D Did 280 300 320 340
 SLR1736 : APPLNTAPITVSHHCTLAATWRRSRDVHLESKTEETASFYQVHWKEEFLEYLYPLTMLPNFSNTIF----- : 308
 SLR0926 : LMLNPVYVMSATATVAGVWVQYIQLSAPTEPKLYGQ--IFGQVYIIGFVILAGSLGWV----- : 292
 SLL1899 : LHOLGILLYATATITGGGFFVKAWOLKQAPGDRDEARG-LPKFSSEYLYMLCLAMVIDSVPVTHQLVVAQMGTLLG : 316
 SLR0056 : YVHQOQYATVLLTETIPPTFDQMYFLRNLLENDAKYQ-ASAQFLVFGVATETALCHAGI----- : 324
 SLR1518 : QAPWQTLTITASIPWAVQVIRHRVVGQYHDQEQVSNCKFTAVNLHFESGMVMAAGYCWAGUG----- : 307
 N 5D Did 320 340

Figure 21

```

      *      20      *      40      *      60      *      80
ATPT2 : -----MESLLSSSLVSAAGFCWKQNLKHLHSLSEIRVLRCDSSKVVAKPKFRNNLVRPDGQSSLLLPKHKSRRFRVNATAG : 80
SLR1736 : -----
ATPT3 : MAFFGLSRVRRLLKSSVSVPSSSSALLQSQHKSLSNPVTHYTNPFCKYPSMNDNYQVMSKRELHQEKFFGVGMVYRLICGMSSS : 89
SLR0926 : -----
ATPT4 : -----WMRRSVVYRFSRISVSSSLPNRLIPMSRELCAVNSFQP-----PVSTESTAKLIGITGRSDANRVFATA : 67
SLL1899 : -----
ATPT12 : -----MTSILNTVSTIHSRVTSDVRGVLSLRNSUSVEFT-----RRRSGFSTLIYESPGRRFVRAETDT : 63
SLR0056 : -----
ATPT8 : -----
SLR1518 : -----MVLTA-----MVLTA-----MTES : 4
      : -----MTES : 4

      *      100      *      120      *      140      *      160      *      1
ATPT2 : PEAFDSNSKOK-----SFRDS-----DAFYR-----FSRPH-----TGTVLSALS-----VSFLAVEKVS-----DISPLFTGILE : 140
SLR1736 : -----MATHQAFWR-----FSRPH-----TGTVLSALS-----VSFLAVEKVS-----DISPLFTGILE : 49
ATPT3 : SSVLEGKPKKDDKEKSDGVVVKKASWIDLYLPEEVRYAKLARLDKPICTWILAWPCWMS-----IALADPGS-----LPSFKYMALFGC : 170
SLR0926 : -----MVAQTPSSP-----PLWITIIYL-----LRMKPAGRIILMIPALWA-----VCLAAQ-----G-----LPPLPUGTIAL : 56
ATPT4 : TAATATATTG-----EISSRVAALAGLGHYAR-----CYWELSKAKLSMLVATSG-----TGILGTGNAAISFPGLCYTCAG : 138
SLL1899 : TKIHRHDSWG-----AVCKSYQOLTTP-----RIIPLLTPTTAAAMMI-----ASEGR-----VDLPKDUITILG : 60
ATPT12 : DKVKSOTPKRAP-----AGSSINQLLGHKAS-----OETNKKIRIQOLTTPVTWPLVMGVVCGAAASGNFHWTPEDVAKSELIC : 139
SLR0056 : QNT-GQNOAKA-----ROLLGKGAAP-----GESSIWKIRIOLMKPITWPLWGVVCGAAASGYIWSVEDFEKALTC : 73
ATPT8 : EVPKLASAAEY-----FFKRGVQKQF-----RSTILLKATALNVRVP-----EALIGEST-----DIVTSELRVRQR : 63
SLR1518 : SPLAPSTAPAT-----RKLWDAAIKP-----PMYAVAVVPITVG-----SAVYGLTG-----QWGHGDVFTIFLL : 59

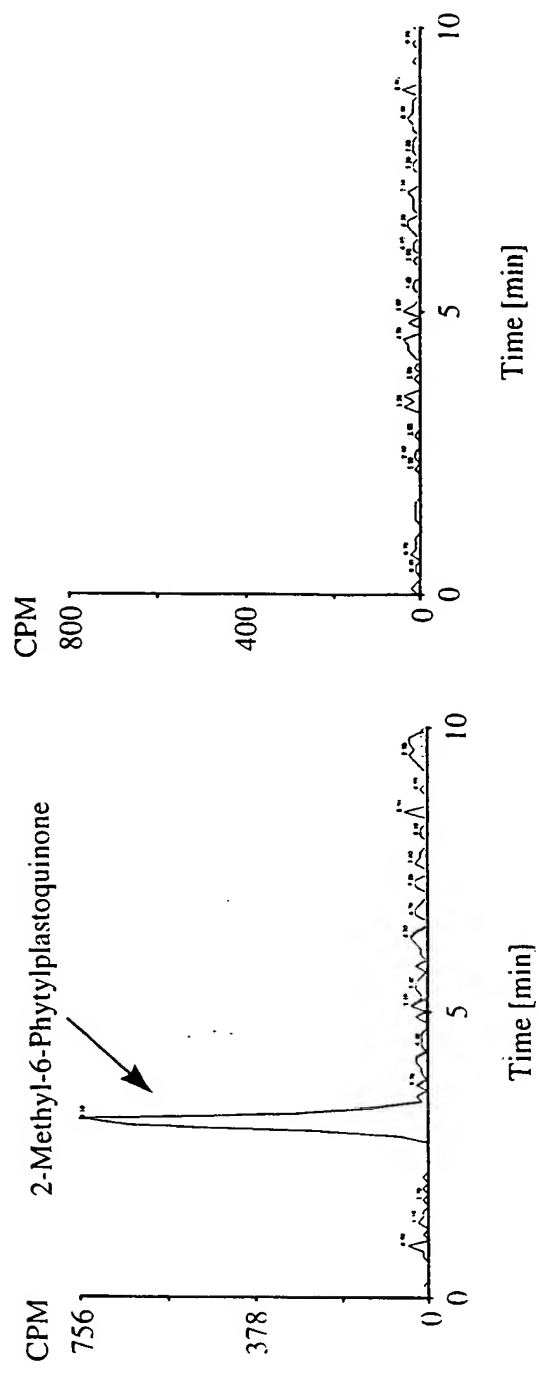
      *      200      *      220      *      240      *      260
ATPT2 : AVVAALMMNIYIVGLNQLSVEIDKVNKPYPPLASGEYSVNTGIAIVASFSMSFMLGWISWPLFWALFVSFMLGTAYS-INPFLR : 228
SLR1736 : AWIACLLGNVYIVGLNQLSVEIDKVNKPYPPLASGEYSVNTGIAIVASFSMSFMLGWISWPLFWALFVSFMLGTAYS-INPFLR : 134
ATPT3 : GAIL-----IRGAGCTINDLDQDITTKVDTKLRPIASGLT-----PFOGEGFDGQOLHGLG-----ILLQNNYSRWLGAS-----SFLVIV : 246
SLR0926 : GTJA-----TSLGCVVADLMDRDPQVETKORLAARHS-VOVGEVGVAVALLICAG-----LFFYLTPTSEFWICVF-----AVPPIV : 132
ATPT4 : TMMI-----AASANSUNOIEISNSKMKRMTMLRLPSGRRESVPHAFATWATAGASGACL-----LSKTNMMAAGLASAN-----LNYAF : 215
SLL1899 : GTJA-----AASAOITLNCIDQDIDYEMLRTRARIPAGKVQPRHAFALAGVTSFAL-----LSTFVNVLGGCLALSG-----IVFYML : 137
ATPT12 : MMLSGPCTGTGYTOTINDWYRRTDAINEPYRPIESGAISEPEVITQWVLTGGGIGILD-VWQGHDPPTMMITLGLG-----GSLFSY : 223
SLR0056 : MMLSGPCTGTGYTOTINDWYRRTDAINEPYRPIESGAISEPEVITQWVLTGGGIGILD-VWQGHDPPTMMITLGLG-----GSLFSY : 157
ATPT8 : GIAE-----ITEMIHVASLHDDVDDADTDRGVGSLNVNMGKMSVAGDFISRACQAL-----ALLKNTFVVALATAVEHLVTGETM : 144
SLR1518 : SATJA-----ITAWINLSQDVFTSDTGIDVRKAHSVNVNLTGNRNVLFLSNFTFLAGVLGLMSMS-----WRAQDWTULELIGVA : 138
      : n d d

```

Figure 22 1/2

[illegible]

Figure 22 2/2



Synechocystis 6803 wild type *Synechocystis* slr1736 knockout

Figure 23

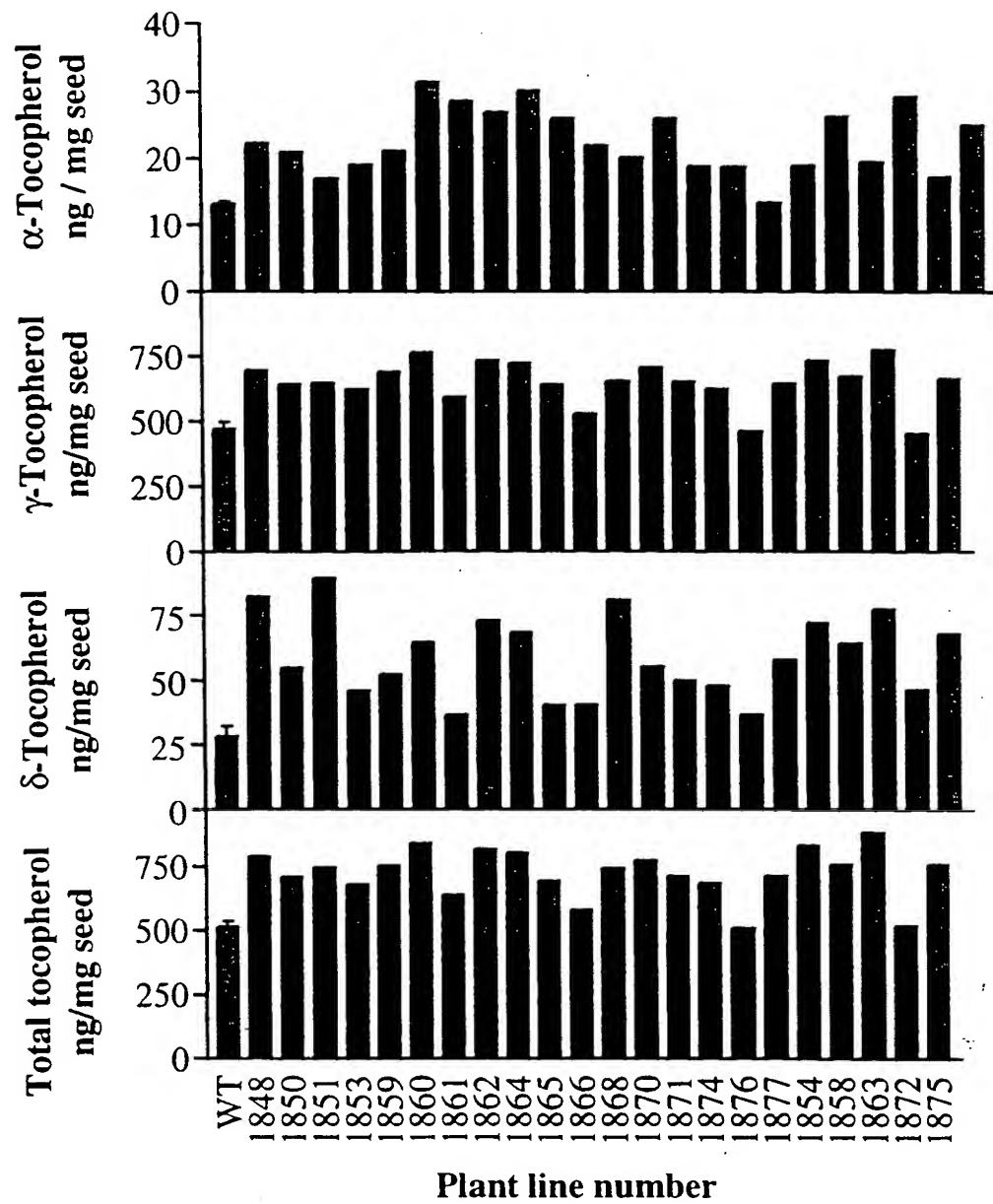


Figure 24

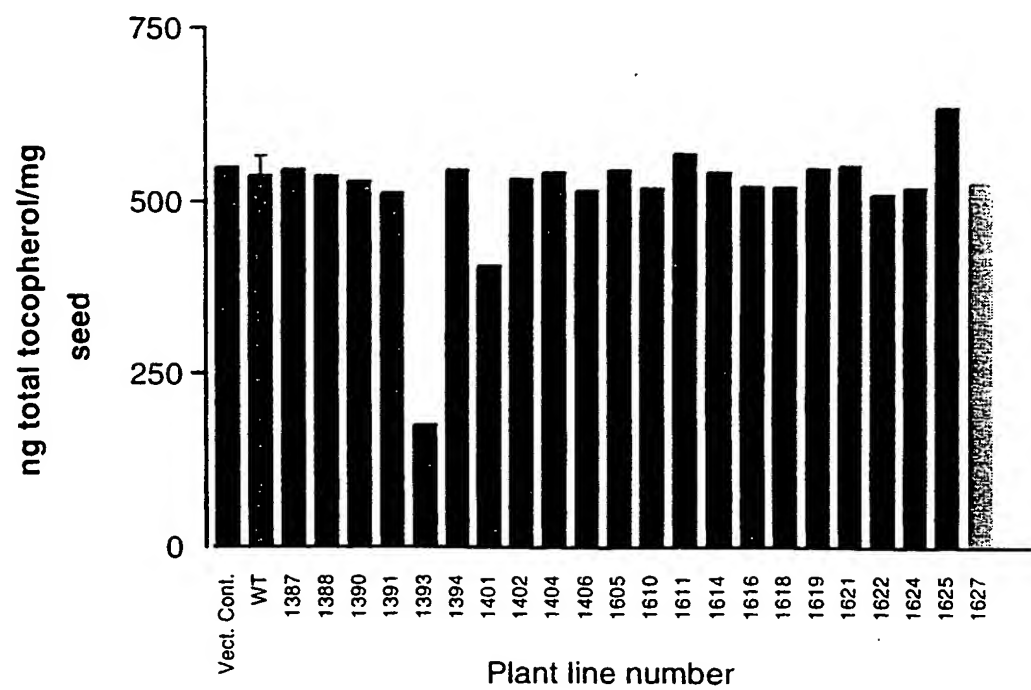


Figure 25

SEQUENCE LISTING

5 <110> Calgene LLC
 <120> Nucleic Acid Sequences Involved in
 Tocopherol Synthesis
 <130> 17133/00/WO
 10 <150> 60/129,899
 <151> 1999-04-15
 <150> 60/146,461
 15 <151> 1999-07-30
 <160> 94
 <170> FastSEQ for Windows Version 4.0
 20 <210> 1
 <211> 1182
 <212> DNA
 <213> Arabidopsis sp
 25 <400> 1
 atggagtcctc tgcctctctag tctctctctt gtttccgctg ctgggtgggtt ttgttggaag 60
 aagcagaatc taaagctcca ctctttatca gaaatccgag ttctgcgttg tgattcgagt 120
 aaagtgtgctg caaaaccgaa gtttaggaac aatcttggtta ggcctgatgg tcaaggatct 180
 30 tcattgttgt tgtatccaaa acataagtcg agatttcggg ttaatgccac tgcgggtcag 240
 cctgaggctt tcgactcgaa tagcaaacag aagtctttta gagactcggt agatgcgttt 300
 tacaggtttt ctaggcctca tacagttatt ggcacagtgc ttagcatttt atctgtatct 360
 ttcttagcag tagagaaggt tcttgatata tctcctttac ttttcactgg catcttgag 420
 gctgttggtg cagctctcat gatgaacatt tacatagttg ggctaaatca gttgtctgat 480
 35 gttgaaatag ataaggttaa caagccctat cttccattgg catcaggaga atattctggt 540
 aacaccggca ttgcaatagt agcttccttc tccatcatga gtttctggct tgggtggatt 600
 gttggttcat ggccattgtt ctgggctctt tttgtgagtt tcatgctcgg tactgcatac 660
 tctatcaatt tgccactttt acgggtggaaa agatttgcatt tgggtgcagc aatgtgtatc 720
 ctgcgtgtcc gagctattat tgttcaaatc gcctttttatc tacatattca gacacatgtg 780
 40 tttggaagac caatcttgtt cactaggcct cttattttctg cactgcgtt tatgagcttt 840
 ttctctgtcg ttattgcatt gtttaaggat atacctgata tcgaagggga taagatatctc 900
 ggaatccgat cattctctgt aactctgggt cagaaacggg tgttttggaac atgtgttaca 960
 ctacttcaaa tggcttacgc tgttgcaatt ctagttaggag ccacatctcc attcatatgg 1020
 agcaaagtca tctcggttgt gggctatgtt atactcgcaa caactttgtg ggctcgagct 1080
 45 aagtcggttg atctgagtag caaaaccgaa ataacttcat gttatatgtt catatggaag 1140
 ctcttttatg cagagtactt gctgttacct tttttgaagt ga 1182

<210> 2
 <211> 393
 <212> PRT
 5 <213> Arabidopsis sp

<400> 2
 Met Glu Ser Leu Leu Ser Ser Ser Ser Leu Val Ser Ala Ala Gly Gly
 1 5 10 15
 10 Phe Cys Trp Lys Lys Gln Asn Leu Lys Leu His Ser Leu Ser Glu Ile
 20 25 30
 Arg Val Leu Arg Cys Asp Ser Ser Lys Val Val Ala Lys Pro Lys Phe
 35 40 45
 Arg Asn Asn Leu Val Arg Pro Asp Gly Gln Gly Ser Ser Leu Leu Leu
 15 50 55 60
 Tyr Pro Lys His Lys Ser Arg Phe Arg Val Asn Ala Thr Ala Gly Gln
 65 70 75 80
 Pro Glu Ala Phe Asp Ser Asn Ser Lys Gln Lys Ser Phe Arg Asp Ser
 85 90 95
 20 Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile Gly Thr
 100 105 110
 Val Leu Ser Ile Leu Ser Val Ser Phe Leu Ala Val Glu Lys Val Ser
 115 120 125
 Asp Ile Ser Pro Leu Leu Phe Thr Gly Ile Leu Glu Ala Val Val Ala
 25 130 135 140
 Ala Leu Met Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp
 145 150 155 160
 Val Glu Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly
 165 170 175
 30 Glu Tyr Ser Val Asn Thr Gly Ile Ala Ile Val Ala Ser Phe Ser Ile
 180 185 190
 Met Ser Phe Trp Leu Gly Trp Ile Val Gly Ser Trp Pro Leu Phe Trp
 195 200 205
 Ala Leu Phe Val Ser Phe Met Leu Gly Thr Ala Tyr Ser Ile Asn Leu
 35 210 215 220
 Pro Leu Leu Arg Trp Lys Arg Phe Ala Leu Val Ala Ala Met Cys Ile
 225 230 235 240
 Leu Ala Val Arg Ala Ile Ile Val Gln Ile Ala Phe Tyr Leu His Ile
 245 250 255
 40 Gln Thr His Val Phe Gly Arg Pro Ile Leu Phe Thr Arg Pro Leu Ile
 260 265 270
 Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala Leu Phe
 275 280 285
 Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser
 45 290 295 300
 Phe Ser Val Thr Leu Gly Gln Lys Arg Val Phe Trp Thr Cys Val Thr

3

Met Ala Phe Phe Gly Leu Ser Arg Val Ser Arg Arg Leu Leu Lys Ser
 1 5 10 15
 Ser Val Ser Val Thr Pro Ser Ser Ser Ser Ala Leu Leu Gln Ser Gln
 20 25 30
 5 His Lys Ser Leu Ser Asn Pro Val Thr Thr His Tyr Thr Asn Pro Phe
 35 40 45
 Thr Lys Cys Tyr Pro Ser Trp Asn Asp Asn Tyr Gln Val Trp Ser Lys
 50 55 60
 Gly Arg Glu Leu His Gln Glu Lys Phe Phe Gly Val Gly Trp Asn Tyr
 10 65 70 75 80
 Arg Leu Ile Cys Gly Met Ser Ser Ser Ser Ser Val Leu Glu Gly Lys
 85 90 95
 Pro Lys Lys Asp Asp Lys Glu Lys Ser Asp Gly Val Val Val Lys Lys
 100 105 110
 15 Ala Ser Trp Ile Asp Leu Tyr Leu Pro Glu Glu Val Arg Gly Tyr Ala
 115 120 125
 Lys Leu Ala Arg Leu Asp Lys Pro Ile Gly Thr Trp Leu Leu Ala Trp
 130 135 140
 Pro Cys Met Trp Ser Ile Ala Leu Ala Ala Asp Pro Gly Ser Leu Pro
 20 145 150 155 160
 Ser Phe Lys Tyr Met Ala Leu Phe Gly Cys Gly Ala Leu Leu Leu Arg
 165 170 175
 Gly Ala Gly Cys Thr Ile Asn Asp Leu Leu Asp Gln Asp Ile Asp Thr
 180 185 190
 25 Lys Val Asp Arg Thr Lys Leu Arg Pro Ile Ala Ser Gly Leu Leu Thr
 195 200 205
 Pro Phe Gln Gly Ile Gly Phe Leu Gly Leu Gln Leu Leu Gly Leu
 210 215 220
 Gly Ile Leu Leu Gln Leu Asn Asn Tyr Ser Arg Val Leu Gly Ala Ser
 30 225 230 235 240
 Ser Leu Leu Leu Val Phe Ser Tyr Pro Leu Met Lys Arg Phe Thr Phe
 245 250 255
 Trp Pro Gln Ala Phe Leu Gly Leu Thr Ile Asn Trp Gly Ala Leu Leu
 260 265 270
 35 Gly Trp Thr Ala Val Lys Gly Ser Ile Ala Pro Ser Ile Val Leu Pro
 275 280 285
 Leu Tyr Leu Ser Gly Val Cys Trp Thr Leu Val Tyr Asp Thr Ile Tyr
 290 295 300
 Ala His Gln Asp Lys Glu Asp Asp Val Lys Val Gly Val Lys Ser Thr
 40 305 310 315 320
 Ala Leu Arg Phe Gly Asp Asn Thr Lys Leu Trp Leu Thr Gly Phe Gly
 325 330 335
 Thr Ala Ser Ile Gly Phe Leu Ala Leu Ser Gly Phe Ser Ala Asp Leu
 340 345 350
 45 Gly Trp Gln Tyr Tyr Ala Ser Leu Ala Ala Ala Ser Gly Gln Leu Gly
 355 360 365

Trp Gln Ile Gly Thr Ala Asp Leu Ser Ser Gly Ala Asp Cys Ser Arg
 370 375 380
 Lys Phe Val Ser Asn Lys Trp Phe Gly Ala Ile Ile Phe Ser Gly Val
 385 390 395 400
 5 Val Leu Gly Arg Ser Phe Gln
 405

<210> 5

<211> 1296

10 <212> DNA

<213> Arabidopsis sp

<400> 5

atgtggcgaa gatctgttgt ttctcgttta tcttcaagaa tctctgtttc ttcttcggtta 60
 15 ccaaacccta gactgattcc ttggtccgc gaattatgtg ccgttaatag cttctcccag 120
 cctccggtct cgacggaatc aactgctaag ttagggatca ctggtgttag atctgatgcc 180
 aatcgagttt ttgccactgc tactgccgcc gctacagcta cagctaccac cggtagagatt 240
 tcgtctagag ttgcggcttt ggctggatta gggcatcact acgctcgttg ttattgggag 300
 ctttctaaag ctaaacttag tatgcttgty gttgcaactt ctggaactgg gtatatcttg 360
 20 ggtacgggaa atgctgcaat tagcttcccg gggctttgtt acacatgtgc aggaaccatg 420
 atgattgctg catctgctaa ttcttgaat cagatttttg agataagcaa tgattctaag 480
 atgaaaagaa cgatgctaag gccattgcct tcaggacgta ttagtgttcc acacgctgtt 540
 gcattgggcta ctattgctgg tgcttctggt gcttgtttgt tggccagcaa gactaatatg 600
 ttggctgctg gacttgcatc tgccaatctt gtactttatg cgtttgttta tactccgttg 660
 25 aagcaacttc accctatcaa tacatgggtt ggcgctgttg ttggtgctat cccacccttg 720
 cttgggtggg cggcagcgtc tggtcagatt tcatacaatt cgatgattct tccagctgct 780
 ctttactttt ggcagatacc tcattttatg gcccttgcac atctctgccg caatgattat 840
 gcagctggag gttacaagat gttgtcactc tttgatccgt cagggaagag aatagcagca 900
 gtggctctaa ggaactgctt ttacatgac cctctcggtt tcatcgcccta tgactggggg 960
 30 ttaacctcaa gttgggtttg cctcgaatca acacttctca cactagcaat cgctgcaaca 1020
 gcattttcat tctaccgaga ccggaccatg cataaagcaa ggaaaatgtt ccatgccagt 1080
 cttctcttcc ttctgtttt catgtctggt cttcttctac accgtgtctc taatgataat 1140
 cagcaacaac tcgtagaaga agccggatta acaaatctct tatctggtga agtcaaaact 1200
 cagaggcgaa agaaacgtgt ggctcaacct ccggtggctt atgcctctgc tgcaccgttt 1260
 35 cctttcctcc cagctccttc cttctactct ccatga 1296

<210> 6

<211> 431

<212> PRT

40 <213> Arabidopsis sp

<400> 6

Met Trp Arg Arg Ser Val Val Tyr Arg Phe Ser Ser Arg Ile Ser Val
 1 5 10 15
 45 Ser Ser Ser Leu Pro Asn Pro Arg Leu Ile Pro Trp Ser Arg Glu Leu
 20 25 30

Cys Ala Val Asn Ser Phe Ser Gln Pro Pro Val Ser Thr Glu Ser Thr
 35 40 45
 Ala Lys Leu Gly Ile Thr Gly Val Arg Ser Asp Ala Asn Arg Val Phe
 50 55 60
 5 Ala Thr Ala Thr Ala Ala Ala Thr Ala Thr Ala Thr Thr Gly Glu Ile
 65 70 75 80
 Ser Ser Arg Val Ala Ala Leu Ala Gly Leu Gly His His Tyr Ala Arg
 85 90 95
 Cys Tyr Trp Glu Leu Ser Lys Ala Lys Leu Ser Met Leu Val Val Ala
 10 100 105 110
 Thr Ser Gly Thr Gly Tyr Ile Leu Gly Thr Gly Asn Ala Ala Ile Ser
 115 120 125
 Phe Pro Gly Leu Cys Tyr Thr Cys Ala Gly Thr Met Met Ile Ala Ala
 130 135 140
 15 Ser Ala Asn Ser Leu Asn Gln Ile Phe Glu Ile Ser Asn Asp Ser Lys
 145 150 155 160
 Met Lys Arg Thr Met Leu Arg Pro Leu Pro Ser Gly Arg Ile Ser Val
 165 170 175
 Pro His Ala Val Ala Trp Ala Thr Ile Ala Gly Ala Ser Gly Ala Cys
 20 180 185 190
 Leu Leu Ala Ser Lys Thr Asn Met Leu Ala Ala Gly Leu Ala Ser Ala
 195 200 205
 Asn Leu Val Leu Tyr Ala Phe Val Tyr Thr Pro Leu Lys Gln Leu His
 210 215 220
 25 Pro Ile Asn Thr Trp Val Gly Ala Val Val Gly Ala Ile Pro Pro Leu
 225 230 235 240
 Leu Gly Trp Ala Ala Ala Ser Gly Gln Ile Ser Tyr Asn Ser Met Ile
 245 250 255
 Leu Pro Ala Ala Leu Tyr Phe Trp Gln Ile Pro His Phe Met Ala Leu
 30 260 265 270
 Ala His Leu Cys Arg Asn Asp Tyr Ala Ala Gly Gly Tyr Lys Met Leu
 275 280 285
 Ser Leu Phe Asp Pro Ser Gly Lys Arg Ile Ala Ala Val Ala Leu Arg
 290 295 300
 35 Asn Cys Phe Tyr Met Ile Pro Leu Gly Phe Ile Ala Tyr Asp Trp Gly
 305 310 315 320
 Leu Thr Ser Ser Trp Phe Cys Leu Glu Ser Thr Leu Leu Thr Leu Ala
 325 330 335
 Ile Ala Ala Thr Ala Phe Ser Phe Tyr Arg Asp Arg Thr Met His Lys
 40 340 345 350
 Ala Arg Lys Met Phe His Ala Ser Leu Leu Phe Leu Pro Val Phe Met
 355 360 365
 Ser Gly Leu Leu Leu His Arg Val Ser Asn Asp Asn Gln Gln Gln Leu
 370 375 380
 45 Val Glu Glu Ala Gly Leu Thr Asn Ser Val Ser Gly Glu Val Lys Thr
 385 390 395 400

Gln Arg Arg Lys Lys Arg Val Ala Gln Pro Pro Val Ala Tyr Ala Ser
 405 410 415
 Ala Ala Pro Phe Pro Phe Leu Pro Ala Pro Ser Phe Tyr Ser Pro
 420 425 430

5
 <210> 7
 <211> 479
 <212> DNA
 <213> Arabidopsis sp

10
 <400> 7
 ggaaactccc ggagcacctg ttgacaggta ccgctaacct taatcgataa tttattttctc 60
 ttgtcaggaa ttatgtaagt ctggtggaag gctcgcatac catttttgca ttgcctttcg 120
 ctatgatcgg gtttactttg ggtgtgatga gaccaggcgt ggcttttatgg tatggcgaaa 180
 15 acccattttt atccaatgct gcattccctc ccgatgattc gttctttcat tcctatacag 240
 gtatcatgct gataaaactg ttactgggtac tggtttgatg ggtatcagca agaagcgcg 300
 cgatggcggt taaccgggtat ctgcacaggc attttgacgc gaagaaccgc cgtactgcc 360
 tccgtgaaat acctgcgggc gtcataatctg ccaacagtgc gctggtgttt acgataggct 420
 gctgcgtggt attctgggtg gcctgttatt tcattaacac gatctgtttt tacctggcg 479

20
 <210> 8
 <211> 551
 <212> DNA
 <213> Arabidopsis sp

25
 <220>
 <221> misc_feature
 <222> (1)...(551)
 <223> n = A,T,C or G

30
 <400> 8
 ttgtggctta caccttaatg agcatagccc agnccattac ggctcgtaa tcggcgccat 60
 ngccgngcgt gntgcaccgg tagtgggcta ctgcgccgtg accaatcagc ttgatctagc 120
 ggctctttatt ctgtttttta ttttactgtt ctggcaaatg ccgcattttt acgcgatttc 180
 35 catttttcagg ctaaaagact tttcagcggc ctgtattccg gtgctgcccc tcattaaaga 240
 cctgcgctat accaaaatca gcatgctggt ttacgtgggc ttatttacac tggctgctat 300
 catgccggcc ctcttagggt atgccggtg gatttatggg atagcggcct taattttagg 360
 cttgtattgg ctttatattg ccatacaagg attcaagacc gccgatgac aaaaatggtc 420
 tcgtaagatg tttggatctt cgattttaat cattaccctc ttgtcggtaa tgatgcttgt 480
 40 ttaaaacttac tgcctcctga agtttatata tcgataattt cagcttaagg aggcttagtg 540
 gttaattcaa t 551

<210> 9
 <211> 297
 <212> PRT
 <213> Arabidopsis sp

<400> 9
 Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
 1 5 10 15
 5 Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
 20 25 30
 Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
 35 40 45
 Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
 10 50 55 60
 Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
 65 70 75 80
 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
 85 90 95
 15 Val Val Met Gly Asn Lys Val Val Ala Leu Leu Ala Thr Ala Val Glu
 100 105 110
 His Leu Val Thr Gly Glu Thr Met Glu Ile Thr Ser Ser Thr Glu Gln
 115 120 125
 Arg Tyr Ser Met Asp Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala
 20 130 135 140
 Ser Leu Ile Ser Asn Ser Cys Lys Ala Val Ala Val Leu Thr Gly Gln
 145 150 155 160
 Thr Ala Glu Val Ala Val Leu Ala Phe Glu Tyr Gly Arg Asn Leu Gly
 165 170 175
 25 Leu Ala Phe Gln Leu Ile Asp Asp Ile Leu Asp Phe Thr Gly Thr Ser
 180 185 190
 Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp Ile Arg His Gly Val Ile
 195 200 205
 Thr Ala Pro Ile Leu Phe Ala Met Glu Glu Phe Pro Gln Leu Arg Glu
 30 210 215 220
 Val Val Asp Gln Val Glu Lys Asp Pro Arg Asn Val Asp Ile Ala Leu
 225 230 235 240
 Glu Tyr Leu Gly Lys Ser Lys Gly Ile Gln Arg Ala Arg Glu Leu Ala
 245 250 255
 35 Met Glu His Ala Asn Leu Ala Ala Ala Ile Gly Ser Leu Pro Glu
 260 265 270
 Thr Asp Asn Glu Asp Val Lys Arg Ser Arg Arg Ala Leu Ile Asp Leu
 275 280 285
 Thr His Arg Val Ile Thr Arg Asn Lys
 40 290 295

 <210> 10
 <211> 561
 <212> DNA
 45 <213> Arabidopsis sp

<400> 10
aagcgcaccc gtcctcttct acgattgccg ccagccgcac gtatggctgc ataaccgacc 60
gccccatccc gctcgcggcc gcggctgaat tcattcacac cgcgacgctg ctgcatgacg 120
acgtcgtcga tgaaagcgat ttgcgccgcg gccgcgaaag cgcgcataag gttttcggca 180
5 atcaggcgag cgtgctcgtc ggcgatttcc tttctctccc cgccttccag ctgatgggtg 240
aagacggctc gctcgcgcg ctgcgcattc tctcggatgc ctccgccgtg atcgcgcagg 300
gcgaagtgat gcagctcggc accgcgcgca atcttgaaac caatatgagc cagtatctcg 360
atgtgatcag cgcgaagacc gccgcgctct ttgccgcgcg ctgcgaaatc ggcccgggtga 420
tggcgaacgc gaaggcggaa gatgctgccg cgatgtgcga atacggcatg aatctcggta 480
10 tcgccttcca gatcatcgac gaccttctcg attacggcac cggcggccac gccgagcttg 540
gcaagaacac gggcgacgat t 561

<210> 11
<211> 966
15 <212> DNA
<213> Arabidopsis sp

<400> 11
atgggtacttg ccgaggttcc aaagcttgcc tctgctgctg agtacttctt caaaaggggt 60
20 gtgcaaggaa aacagtttctg ttcaactatt ttgctgctga tggcgacagc tctgaatgta 120
cgcgttccag aagcattgat tggggaatca acagatatag tcacatcaga attacgcgta 180
aggcaacggg gtattgctga aatcactgaa atgatacacg tcgcaagtct actgcacgat 240
gatgtcttgg atgatgccga tacaaggcgt ggtgttggtt ccttaaatgt tgtaatgggt 300
aacaagatgt cggattatgc aggagacttc ttgctctccc gggcttgggt ggctctcgct 360
25 gctttaaaga acacagaggt ttagcatta cttgcaactg ctgtagaaca tcttgttacc 420
ggtgaaacca tggaaataac tagttcaacc gagcagcgtt atagtatgga ctactacatg 480
cagaagacat attataagac agcatcgcta atctctaaca gctgcaaagc tgttgccgtt 540
ctcactggac aaacagcaga agttgccgtg ttacttttgg agtatgggag gaatctgggt 600
ttagcattcc aattaataga cgacattctt gatttcacgg gcacatctgc ctctctcgga 660
30 aagggatcgt tgtcagatat tcgccatgga gtcataacag cccaatcct ctttgccatg 720
gaagagtttc ctcaactacg cgaagttggt gatcaagttg aaaaagatcc taggaatggt 780
gacattgctt tagagtatct tgggaagagc aagggaatac agagggcaag agaattagcc 840
atggaacatg cgaatctagc agcagctgca atcgggtctc tacctgaaac agacaatgaa 900
gatgtcaaaa gatcgaggcg ggcacttatt gacttgaccc atagagtcac caccagaaac 960
35 aagtga 966

<210> 12
<211> 321
<212> PRT
40 <213> Arabidopsis sp

<400> 12
Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
1 5 10 15
45 Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
20 25 30

Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
 35 40 45
 Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
 50 55 60
 5 Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
 65 70 75 80
 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
 85 90 95
 Val Val Met Gly Asn Lys Met Ser Val Leu Ala Gly Asp Phe Leu Leu
 10 100 105 110
 Ser Arg Ala Cys Gly Ala Leu Ala Ala Leu Lys Asn Thr Glu Val Val
 115 120 125
 Ala Leu Leu Ala Thr Ala Val Glu His Leu Val Thr Gly Glu Thr Met
 130 135 140
 15 Glu Ile Thr Ser Ser Thr Glu Gln Arg Tyr Ser Met Asp Tyr Tyr Met
 145 150 155 160
 Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys
 165 170 175
 Ala Val Ala Val Leu Thr Gly Gln Thr Ala Glu Val Ala Val Leu Ala
 20 180 185 190
 Phe Glu Tyr Gly Arg Asn Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp
 195 200 205
 Ile Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu
 210 215 220
 25 Ser Asp Ile Arg His Gly Val Ile Thr Ala Pro Ile Leu Phe Ala Met
 225 230 235 240
 Glu Glu Phe Pro Gln Leu Arg Glu Val Val Asp Gln Val Glu Lys Asp
 245 250 255
 Pro Arg Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys Ser Lys Gly
 30 260 265 270
 Ile Gln Arg Ala Arg Glu Leu Ala Met Glu His Ala Asn Leu Ala Ala
 275 280 285
 Ala Ala Ile Gly Ser Leu Pro Glu Thr Asp Asn Glu Asp Val Lys Arg
 290 295 300
 35 Ser Arg Arg Ala Leu Ile Asp Leu Thr His Arg Val Ile Thr Arg Asn
 305 310 315 320
 Lys

40 <210> 13
 <211> 621
 <212> DNA
 <213> Arabidopsis sp

45 <400> 13
 gctttctcct ttgctaattc ttgagctttc ttgatccac cgcgatttct aactatttca 60

atcgcttctt caagcgatcc aggctcacia aactcagact caatgatctc tcttagcctt 120
 ggctcattct ctacgcgcaa gatcactggc gccgttatgt tacctttggc taagtcatta 180
 gctgcaggct tacctaactg ctctgtggac tgagtgaagt ccagaatgac atcaactact 240
 tgaagagata aaccgagatt cttcccgaac tgatacattt gctctgcgac cttgctttcg 300
 5 actttactga aaattgctgc tcctttgggtg cttgcagcta ctaatgaagc tgtctttag 360
 taactcttta gcatgtagtc atcaagcttg acatcacaaat cgaataaact cgatgcttgc 420
 tttatctcac cgcttgcaaa atctttgatc acctgcaaaa agataaatca agattcagac 480
 caaatgttct ttgtattgag tagcttcatc taatctcaga aaggaatatt acctgactta 540
 tgagcttaat gacttcaagg ttttcgagat ttgtaagtac catgatgctt gagcaacatg 600
 10 aaatccccag ctaatacagc t 621

<210> 14

<211> 741

<212> DNA

15 <213> Arabidopsis sp

<400> 14

ggtgagtttt gttaatagtt atgagattca tctatttttg tcataaaatt gtttggtttg 60
 gtttaaactc tgtgtataat tgcaggaaag gaaacagttc atgagctttt cggcacaaaga 120
 20 gtagcgggtgc tagctggaga tttcatgttt gctcaagcgt catggtaact agcaaatctc 180
 gagaatcttg aagttattaa gctcatcagt caggtaactta gttactctta cattgttttt 240
 ctatgagggt gagctatgaa tctcatttcg ttgaataatg ctgtgcctca aacttttttt 300
 catgttttca ggtgatcaaa gactttgcaa gcggagagat aaagcaggcg tccagcttat 360
 ttgactgcga caccaagctc gacgagtact tactcaaaag tttctacaag acagcctctt 420
 25 tagtggctgc gagcaccaaa ggagctgcca ttttcagcag agttgagcct gatgtgacag 480
 aacaaatgta cgagtttggg aagaatctcg gtctctcttt ccagatagtt gatgataatt 540
 tggattttcac tcagtcgaca gagcagctcg ggaagccagc agggagtgat ttggctaaag 600
 gtaacttaac agcacctgtg attttcgctc tggagaggga gccaaaggcta agagagatca 660
 ttgagtcaaa gttctgtgag gcgggttctc tggagaagc gattgaagcg gtgacaaaag 720
 30 gtggggggat taagagagca c 741

<210> 15

<211> 1087

<212> DNA

35 <213> Arabidopsis sp

<400> 15

cctcttcagc caatccagag gaagaagaga caacttttta tctttcgtca agagtctccg 60
 aaaacgcacg gttttatgct ctctctctcg ccctcacctc acaagacgca gggcacatga 120
 40 ttcaaccaga gggaaaaagc aacgataaca actctgcttt tgatttcaag ctgtatatga 180
 tccgcaaagc cgagtctgta aatgcggctc tcgacgtttc cgtaccgctt ctgaaacccc 240
 ttacgatcca agaagcggtc aggtactctt tgctagccgg cggaaaacgt gtgaggcctc 300
 tgctctgcat tgccgcttgt gagcttgtgg ggggagacga ggctactgcc atgtcagccg 360
 cttgcgcggc cgagatgatc cacacaagct ctctcattca tgacgatctt ccgtgcatgg 420
 45 acaatgccga cctccgtaga ggcaagccca ccaatcacia ggtatgttgt ttaattatat 480
 gaaggctcag agataatgct gaactagtgt tgaaccaatt tttgctcaaa caaggtatat 540

ggagaagaca tggcgggtttt ggcagggtgat gcactccttg cattggcggtt tgagcacatg 600
 acggtttgtt cgagtgggtt ggtcgctccc gagaagatga ttcgcgccgt gggttgagctg 660
 gccaggggcca tagggactac agggctagtt gctggacaaa tgatagacct agccagcgaa 720
 agactgaatc cagacaagggt tggattggag catctagagt tcatccatct ccacaaaacg 780
 5 ggcggcattgt tggaggcagc ggcagtttta ggggttataa tgggaggttg aacagaggaa 840
 gaaatcgaaa agcttagaaa gtatgctagg tgtattggac tactgtttca gggtgttgat 900
 gacattctcg acgtaacaaa atctactgag gaattgggta agacagccgg aaaagacgta 960
 atggccggaa agctgacgta tccaaggctg ataggttttg agggatccag ggaagttgca 1020
 gagcacctga ggagagaagc agaggaaaag cttaaagggt ttgatccaag tcaggcggcg 1080
 10 cctctgg 1087

<210> 16

<211> 1164

<212> DNA

15 <213> Arabidopsis sp

<400> 16

atgacttcga ttctcaacac tgtctccacc atccactctt ccagagttac ctccgtcgat 60
 cgagtcggag tctctctctt tcggaattcg gattccgttg agttcactcg ccggcggttct 120
 20 ggtttctcga cgttgatcta cgaatcaccc gggcggagat ttgttggtcg tgcggcggag 180
 actgatactg ataaagttaa atctcagaca cctgacaagg caccagccgg tggttcaagc 240
 attaaccagc ttctcggtat caaaggagca tctcaagaaa ctaataaatg gaagattcgt 300
 cttcagctta caaaaccagt cacttggcct ccactgggtt ggggagtcgt ctgtggtgct 360
 gctgcttcag ggaactttca ttggacccca gaggatgttg ctaagtcgat tctttgcatg 420
 25 atgatgtctg gtccttgtct tactggctat acacagacaa tcaacgactg gtatgataga 480
 gatatcgacg caattaatga gccatatact ccaattccat ctggagcaat atcagagcca 540
 gaggttatta cacaagtctg ggtgctatta ttgggaggtc ttggtattgc tggaaatatta 600
 gatgtgtggg cagggcatac cactcccact gtcttctatc ttgctttggg aggatcattg 660
 ctatcttata tatactctgc tccacctctt aagctaaaac aaaatggatg ggttggaat 720
 30 tttgcacttg gagcaagcta tattagtttg ccatgggtggg ctggccaagc attgtttggc 780
 actcttacgc cagatgttgt tgttctaaca ctctgtaca gcatagctgg gttaggaata 840
 gccattgtta acgacttcaa aagtgttgaa ggagatagag cattaggact tcagtctctc 900
 ccagtagctt ttggcaccga aactgcaaaa tggatatgctg ttggtgctat agacattact 960
 cagctttctg ttgccggata tctattagca tctgggaaac cttattatgc gttggcggtg 1020
 35 gttgctttga tcattctca gattgtgttc cagtttaaact actttctcaa ggacctgtc 1080
 aaatacgacg tcaagtacca ggcaagcgcg cagccattct tgggtgctcg aatatttgta 1140
 acggcattag catcgcaaca ctga 1164

<210> 17

40 <211> 387

<212> PRT

<213> Arabidopsis sp

<400> 17

45 Met Thr Ser Ile Leu Asn Thr Val Ser Thr Ile His Ser Ser Arg Val
 1 5 10 15

Thr Ser Val Asp Arg Val Gly Val Leu Ser Leu Arg Asn Ser Asp Ser
 20 25 30
 Val Glu Phe Thr Arg Arg Arg Ser Gly Phe Ser Thr Leu Ile Tyr Glu
 35 40 45
 5 Ser Pro Gly Arg Arg Phe Val Val Arg Ala Ala Glu Thr Asp Thr Asp
 50 55 60
 Lys Val Lys Ser Gln Thr Pro Asp Lys Ala Pro Ala Gly Gly Ser Ser
 65 70 75 80
 Ile Asn Gln Leu Leu Gly Ile Lys Gly Ala Ser Gln Glu Thr Asn Lys
 10 85 90 95
 Trp Lys Ile Arg Leu Gln Leu Thr Lys Pro Val Thr Trp Pro Pro Leu
 100 105 110
 Val Trp Gly Val Val Cys Gly Ala Ala Ala Ser Gly Asn Phe His Trp
 115 120 125
 15 Thr Pro Glu Asp Val Ala Lys Ser Ile Leu Cys Met Met Met Ser Gly
 130 135 140
 Pro Cys Leu Thr Gly Tyr Thr Gln Thr Ile Asn Asp Trp Tyr Asp Arg
 145 150 155 160
 Asp Ile Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala
 20 165 170 175
 Ile Ser Glu Pro Glu Val Ile Thr Gln Val Trp Val Leu Leu Gly
 180 185 190
 Gly Leu Gly Ile Ala Gly Ile Leu Asp Val Trp Ala Gly His Thr Thr
 195 200 205
 25 Pro Thr Val Phe Tyr Leu Ala Leu Gly Gly Ser Leu Leu Ser Tyr Ile
 210 215 220
 Tyr Ser Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Val Gly Asn
 225 230 235 240
 Phe Ala Leu Gly Ala Ser Tyr Ile Ser Leu Pro Trp Trp Ala Gly Gln
 30 245 250 255
 Ala Leu Phe Gly Thr Leu Thr Pro Asp Val Val Val Leu Thr Leu Leu
 260 265 270
 Tyr Ser Ile Ala Gly Leu Gly Ile Ala Ile Val Asn Asp Phe Lys Ser
 275 280 285
 35 Val Glu Gly Asp Arg Ala Leu Gly Leu Gln Ser Leu Pro Val Ala Phe
 290 295 300
 Gly Thr Glu Thr Ala Lys Trp Ile Cys Val Gly Ala Ile Asp Ile Thr
 305 310 315 320
 Gln Leu Ser Val Ala Gly Tyr Leu Leu Ala Ser Gly Lys Pro Tyr Tyr
 40 325 330 335
 Ala Leu Ala Leu Val Ala Leu Ile Ile Pro Gln Ile Val Phe Gln Phe
 340 345 350
 Lys Tyr Phe Leu Lys Asp Pro Val Lys Tyr Asp Val Lys Tyr Gln Ala
 355 360 365
 45 Ser Ala Gln Pro Phe Leu Val Leu Gly Ile Phe Val Thr Ala Leu Ala
 370 375 380

Ser Gln His
385

5 <210> 18
<211> 981
<212> DNA
<213> Arabidopsis sp

<400> 18
10 atgttggttta gtggttcagc gatcccattha agcagcttct gctctcttcc ggagaaaccc 60
cacactcttc ctatgaaact ctctcccgtc gcaatccgat cttcatcctc atctgccccg 120
gggtcggtga acttcgatct gaggacgtat tggacgactc tgatcaccga gatcaaccag 180
aagctggatg aggccatacc ggtcaagcac cctgcgggga tctacgaggc tatgagatac 240
tctgtactcg cacaaggcgc caagcgtgcc cctcctgtga tgtgtgtggc ggcttgcgag 300
15 ctcttcggtg gcgatcgctt cgccgcttcc cccaccgcct gtgccctaga aatggtgcac 360
gcggcttcgt tgatacacga cgacctcccc tgtatggacg acgatcctgt gcgcagagga 420
aagccatcta accacactgt ctacggctct ggcatggcca ttctcgccgg tgacgcctc 480
ttcccactcg ccttccagca cattgtctcc cacacgcctc ctgacctgtg tccccgagcc 540
accatcctca gactcatcac tgagattgcc cgcactgtcg gctccactgg tatggctgca 600
20 ggccagtacg tcgacctga aggaggtccc ttctctctt cctttgttca ggagaagaaa 660
ttcggagcca tgggtgaatg ctctgccgtg tgcggtggcc tattgggagg tgccactgag 720
gatgagctcc agagtctccg aaggtacggg agagccgtcg ggatgctgta tcagggtggtc 780
gatgacatca ccgaggacaa gaagaagagc tatgatggtg gagcagagaa gggaatgatg 840
gaaatggcgg aagagctcaa ggagaaggcg aagaaggagc ttcaagtgtt tgacaacaag 900
25 tatggaggag gagacacact tgttcctctc tacaccttcg ttgactacgc tgctcatcga 960
cattttcttc ttccccctctg a 981

<210> 19
<211> 245
30 <212> DNA
<213> Glycine sp

<400> 19
35 gcaacatctg ggactgggtt tgtcttgggg agtggtagtg ctgttgatct ttcggcactt 60
tcttgcaactt gcttgggtac catgatggtt gctgcatctg ctaactcttt gaatcagggtg 120
tttgagatca ataagatgc taaaatgaag agaacaagtc gcaggccact accctcagga 180
cgcatacaca tacctcatgc agttggctgg gcatcctctg ttggattagc tggtagcggt 240
ctact 245

40 <210> 20
<211> 253
<212> DNA
<213> Glycine sp

45 <400> 20
attggcttcc caagatcatt gggttttctt gttgcattca tgaccttcta ctccctgggt 60

ttggcattgt ccaaggatat acctgacgtt gaaggagata aagagcacgg cattgattct 120
 tttgcagtac gtctagggtca gaaacgggca ttttggattt gcgtttcctt ttttgaaatg 180
 gctttcggag ttggtatcct ggccggagca tcatgctcac acttttggac taaaattttc 240
 acgggtatgg gaa 253
 5
 <210> 21
 <211> 275
 <212> DNA
 <213> Glycine sp
 10
 <400> 21
 tgatcttcta ctctctgggt atggcattgt ccaaggatat atctgacgtt aaaggagata 60
 aagcatcacg catcgatact ttagcgatac gtttgggtca aaaatgggta ttttggattt 120
 gcattatcct ttttgaaatg gcttttggag ttgccctctt ggcaggagca acatcttctt 180
 15 acctttggat taaaattgtc acgggtctgg gacatgctat tcttgcttca attctcttgt 240
 accaagccaa atctatatac ttgagcaaca aagtt 275
 <210> 22
 <211> 299
 20 <212> DNA
 <213> Glycine sp
 <220>
 <221> misc_feature
 25 <222> (1)...(299)
 <223> n = A,T,C or G
 <400> 22
 ccanaatang tncatcttng aaagacaatt ggcctcttca acacacaagt ctgcatgtga 60
 30 agaagaggcc aattgtcttt ccaagatcac ttatngtggc tattgtaatc atgaacttct 120
 tctttgtggg tatggcattg gcaaaggata tacctanctg ttgaaggaga taaaatatat 180
 ggcattgata cttttgcaat acgtataggt caaaaacaag tattttggat ttgtattttc 240
 ctttttgaaa ggctttcgga gtttccttag tggcaggagc aacatcttct agccttggg 299
 35 <210> 23
 <211> 767
 <212> DNA
 <213> Glycine sp
 40 <400> 23
 gtggaggctg tgggttctgc cctgtttatg aatatttata ttgttgggtt gaatcaattg 60
 tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat 120
 tcctttgaaa ctggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttgge 180
 tgggtttag gttcatggcc attatttttg gccctttttg taagctttgt gctaggaact 240
 45 gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgtc tgcagcgatg 300
 tgcattctag ctgttcgggc agtaatatgt caacttgcac ttttccttca catgcagact 360

```

catgtgtaca agaggccacc tgtcttttca agaccattga tttttgctac tgcattcatg      420
agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa      480
gtatttggca tccaatcttt ttcagtgtgt ttaggtcaga agccgggtgt ctggacttgt      540
gttacccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt      600
5  ctttggagca aaattttcac gggctctggga cacgctgtgc tggtttcaat tctctggttt      660
catgccaaat ctgtagattt gaaaagcaaa gtttcgataa catccttcta tatgtttatt      720
tggaagctat tttatgcaga atacttactc attccttttg ttagatg      767

<210> 24
10 <211> 255
    <212> PRT
    <213> Glycine sp

<400> 24
15 Val Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly
    1          5          10          15
    Leu Asn Gln Leu Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr
        20          25          30
    Leu Pro Leu Ala Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile
20      35          40          45
    Val Ala Ser Phe Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly
        50          55          60
    Ser Trp Pro Leu Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr
        65          70          75          80
25 Ala Tyr Ser Ile Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val
        85          90          95
    Leu Ala Ala Met Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu
        100          105          110
    Ala Phe Phe Leu His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val
30      115          120          125
    Phe Ser Arg Pro Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser
        130          135          140
    Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys
        145          150          155          160
35 Val Phe Gly Ile Gln Ser Phe Ser Val Cys Leu Gly Gln Lys Pro Val
        165          170          175
    Phe Trp Thr Cys Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu
        180          185          190
    Leu Val Gly Ala Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly
40      195          200          205
    Leu Gly His Ala Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser
        210          215          220
    Val Asp Leu Lys Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile
        225          230          235          240
45 Trp Lys Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
        245          250          255

```

<210> 25
 <211> 360
 <212> DNA
 5 <213> Zea sp

<220>
 <221> misc_feature
 <222> (1)...(360)
 10 <223> n = A,T,C or G

<400> 25
 ggcgtcttca cttgttcttg tcttctcgta tcccctgatg aagaggttca ctttttgccc 60
 tcaggcttat cttggcctga cattcaactg gggagcttta ctagggtggg ctgctattaa 120
 15 ggaaagcata gaccctgcaa atcatccttc cattgtatac agctgggtatt tgttggacgc 180
 tgggtgtatga tactatatat gcgcatacagg tgttctcgcta tccctacttt catattaatc 240
 cttgatgaag tggccatttc atgttgcgc ggtggtctta tacttgcata tctccatgca 300
 tctcaggaca aagangatga cctgaaagta ggagtccaag tccacagctt aagatttggg 360

20 <210> 26
 <211> 299
 <212> DNA
 <213> Zea sp

25 <220>
 <221> misc_feature
 <222> (1)...(299)
 <223> n = A,T,C or G

30 <400> 26
 gatggttgca gcatctgcaa ataccctcaa ccagggtgtt gngataaaaa atgatgctaa 60
 aatgaaaagg acaatgcgtg cccctcgcca tctggctcgca ttagtctctgc acatgctgcg 120
 atgtgggcta caagtgttgg agttgcagga acagctttgt tggcctggaa ggctaattggc 180
 ttggcagctg ggcttgcagc ttctaattct gttctgtatg catttgtgta tacgccgttg 240
 35 aagcaaatac accctgttaa tacatgggtt ggggcagtcg ttggtgccat cccaccact 299

<210> 27
 <211> 255
 <212> DNA
 40 <213> Zea sp

<220>
 <221> misc_feature
 <222> (1)...(255)
 45 <223> n = A,T,C or G

<400> 27
 anacttgc atctccatgc ntctcaggac aaagangatg acctgaaagt aggtgtcaag 60
 tccacagcat taagatttgg agatttgacc nnatactgna tcagtggctt tggcgcggca 120
 tgcttcggca gcttagcact cagtgggtac aatgctgacc ttggttggtg tttagtgtga 180
 5 tgcttgagcg aagaatggta tngtttttac ttgatattga ctccagacct gaaatcatgt 240
 tggacagggg ggccc 255

<210> 28
 <211> 257
 10 <212> DNA
 <213> Zea sp

<400> 28
 attgaagggg ataggactct ggggcttcag tcacttcctg ttgcttttgg gatggaaact 60
 15 gcaaaatgga tttgtgttgg agcaattgat atcactcaat tatctgttgc aggttaccta 120
 ttgagcaccg gtaagctgta ttatgccctg gtgttgcttg ggctaacaat tcctcaggtg 180
 ttctttcagt tccagtactt cctgaaggac cctgtgaagt atgatgtcaa atatcaggca 240
 agcgcaaac cattctt 257

20 <210> 29
 <211> 368
 <212> DNA
 <213> Zea sp

25 <400> 29
 atccagttgc aaataataat ggcgttcttc tctgttgtaa tagcactatt caaggatata 60
 cctgacatcg aaggggaccg catattcggg atccgatcct tcagcgctccg gttagggcaa 120
 aagaaggctt tttggatctg cgttggcttg cttgagatgg cctacagcgt tgcgatactg 180
 atgggagcta cctcttcttg tttgtggagc aaaacagcaa ccatcgctgg ccattccata 240
 30 cttgccgcga tcctatggag ctgcgcgcga tcggtggact tgacgagcaa agccgcaata 300
 acgtccttct acatgttcat ctggaagctg ttctacgcgg agtacctgct catcctctg 360
 gtgcggtg 368

<210> 30
 35 <211> 122
 <212> PRT
 <213> Zea sp

<400> 30
 40 Ile Gln Leu Gln Ile Ile Met Ala Phe Phe Ser Val Val Ile Ala Leu
 1 5 10 15
 Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg Ile Phe Gly Ile Arg
 20 25 30
 Ser Phe Ser Val Arg Leu Gly Gln Lys Lys Val Phe Trp Ile Cys Val
 45 35 40 45
 Gly Leu Leu Glu Met Ala Tyr Ser Val Ala Ile Leu Met Gly Ala Thr

```

      50              55              60
Ser Ser Cys Leu Trp Ser Lys Thr Ala Thr Ile Ala Gly His Ser Ile
65              70              75              80
Leu Ala Ala Ile Leu Trp Ser Cys Ala Arg Ser Val Asp Leu Thr Ser
5              85              90              95
Lys Ala Ala Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe Tyr
              100              105              110
Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg
              115              120
10
<210> 31
<211> 278
<212> DNA
<213> Zea sp

15
<400> 31
tattcagcac cacctctcaa gctcaagcag aatggatgga ttgggaactt cgctctgggt      60
gcgagttaca tcagcttgcc ctggtgggct ggccaggcgt tatttggaac tcttacacca      120
gatatcattg tcttgactac tttgtacagc atagctgggc tagggattgc tattgtaaat      180
20 gatttcaaga gtattgaagg ggataggact ctggggcttc agtcacttcc tgttgctttt      240
gggatggaaa ctgcaaaatg gatttgtgtt ggagcaat      278

<210> 32
<211> 292
25 <212> PRT
<213> Synechocystis sp

<400> 32
Met Val Ala Gln Thr Pro Ser Ser Pro Pro Leu Trp Leu Thr Ile Ile
30 1              5              10              15
Tyr Leu Leu Arg Trp His Lys Pro Ala Gly Arg Leu Ile Leu Met Ile
              20              25              30
Pro Ala Leu Trp Ala Val Cys Leu Ala Ala Gln Gly Leu Pro Pro Leu
              35              40              45
35 Pro Leu Leu Gly Thr Ile Ala Leu Gly Thr Leu Ala Thr Ser Gly Leu
              50              55              60
Gly Cys Val Val Asn Asp Leu Trp Asp Arg Asp Ile Asp Pro Gln Val
65              70              75              80
Glu Arg Thr Lys Gln Arg Pro Leu Ala Ala Arg Ala Leu Ser Val Gln
40              85              90              95
Val Gly Ile Gly Val Ala Leu Val Ala Leu Leu Cys Ala Ala Gly Leu
              100              105              110
Ala Phe Tyr Leu Thr Pro Leu Ser Phe Trp Leu Cys Val Ala Ala Val
              115              120              125
45 Pro Val Ile Val Ala Tyr Pro Gly Ala Lys Arg Val Phe Pro Val Pro
              130              135              140

```

Gln Leu Val Leu Ser Ile Ala Trp Gly Phe Ala Val Leu Ile Ser Trp
 145 150 155 160
 Ser Ala Val Thr Gly Asp Leu Thr Asp Ala Thr Trp Val Leu Trp Gly
 165 170 175
 5 Ala Thr Val Phe Trp Thr Leu Gly Phe Asp Thr Val Tyr Ala Met Ala
 180 185 190
 Asp Arg Glu Asp Asp Arg Arg Ile Gly Val Asn Ser Ser Ala Leu Phe
 195 200 205
 Phe Gly Gln Tyr Val Gly Glu Ala Val Gly Ile Phe Phe Ala Leu Thr
 210 215 220
 10 Ile Gly Cys Leu Phe Tyr Leu Gly Met Ile Leu Met Leu Asn Pro Leu
 225 230 235 240
 Tyr Trp Leu Ser Leu Ala Ile Ala Ile Val Gly Trp Val Ile Gln Tyr
 245 250 255
 15 Ile Gln Leu Ser Ala Pro Thr Pro Glu Pro Lys Leu Tyr Gly Gln Ile
 260 265 270
 Phe Gly Gln Asn Val Ile Ile Gly Phe Val Leu Leu Ala Gly Met Leu
 275 280 285
 Leu Gly Trp Leu
 20 290

 <210> 33
 <211> 316
 <212> PRT
 25 <213> Synechocystis sp

 <400> 33
 Met Val Thr Ser Thr Lys Ile His Arg Gln His Asp Ser Met Gly Ala
 1 5 10 15
 30 Val Cys Lys Ser Tyr Tyr Gln Leu Thr Lys Pro Arg Ile Ile Pro Leu
 20 25 30
 Leu Leu Ile Thr Thr Ala Ala Ser Met Trp Ile Ala Ser Glu Gly Arg
 35 40 45
 Val Asp Leu Pro Lys Leu Leu Ile Thr Leu Leu Gly Gly Thr Leu Ala
 50 55 60
 35 Ala Ala Ser Ala Gln Thr Leu Asn Cys Ile Tyr Asp Gln Asp Ile Asp
 65 70 75 80
 Tyr Glu Met Leu Arg Thr Arg Ala Arg Pro Ile Pro Ala Gly Lys Val
 85 90 95
 40 Gln Pro Arg His Ala Leu Ile Phe Ala Leu Ala Leu Gly Val Leu Ser
 100 105 110
 Phe Ala Leu Leu Ala Thr Phe Val Asn Val Leu Ser Gly Cys Leu Ala
 115 120 125
 Leu Ser Gly Ile Val Phe Tyr Met Leu Val Tyr Thr His Trp Leu Lys
 45 130 135 140
 Arg His Thr Ala Gln Asn Ile Val Ile Gly Gly Ala Ala Gly Ser Ile

145 150 155 160
 Pro Pro Leu Val Gly Trp Ala Ala Val Thr Gly Asp Leu Ser Trp Thr
 165 170 175
 Pro Trp Val Leu Phe Ala Leu Ile Phe Leu Trp Thr Pro Pro His Phe
 5 180 185 190
 Trp Ala Leu Ala Leu Met Ile Lys Asp Asp Tyr Ala Gln Val Asn Val
 195 200 205
 Pro Met Leu Pro Val Ile Ala Gly Glu Glu Lys Thr Val Ser Gln Ile
 210 215 220
 10 Trp Tyr Tyr Ser Leu Leu Val Val Pro Phe Ser Leu Leu Leu Val Tyr
 225 230 235 240
 Pro Leu His Gln Leu Gly Ile Leu Tyr Leu Ala Ile Ala Ile Ile Leu
 245 250 255
 Gly Gly Gln Phe Leu Val Lys Ala Trp Gln Leu Lys Gln Ala Pro Gly
 15 260 265 270
 Asp Arg Asp Leu Ala Arg Gly Leu Phe Lys Phe Ser Ile Phe Tyr Leu
 275 280 285
 Met Leu Leu Cys Leu Ala Met Val Ile Asp Ser Leu Pro Val Thr His
 290 295 300
 20 Gln Leu Val Ala Gln Met Gly Thr Leu Leu Leu Gly
 305 310 315

 <210> 34
 <211> 324
 25 <212> PRT
 <213> Synechocystis sp

 <400> 34
 Met Ser Asp Thr Gln Asn Thr Gly Gln Asn Gln Ala Lys Ala Arg Gln
 30 1 5 10 15
 Leu Leu Gly Met Lys Gly Ala Ala Pro Gly Glu Ser Ser Ile Trp Lys
 20 25 30
 Ile Arg Leu Gln Leu Met Lys Pro Ile Thr Trp Ile Pro Leu Ile Trp
 35 40 45
 35 Gly Val Val Cys Gly Ala Ala Ser Ser Gly Gly Tyr Ile Trp Ser Val
 50 55 60
 Glu Asp Phe Leu Lys Ala Leu Thr Cys Met Leu Leu Ser Gly Pro Leu
 65 70 75 80
 Met Thr Gly Tyr Thr Gln Thr Leu Asn Asp Phe Tyr Asp Arg Asp Ile
 40 85 90 95
 Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala Ile Ser
 100 105 110
 Val Pro Gln Val Val Thr Gln Ile Leu Ile Leu Leu Val Ala Gly Ile
 115 120 125
 45 Gly Val Ala Tyr Gly Leu Asp Val Trp Ala Gln His Asp Phe Pro Ile
 130 135 140

Met Met Val Leu Thr Leu Gly Gly Ala Phe Val Ala Tyr Ile Tyr Ser
 145 150 155 160
 Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Leu Gly Asn Tyr Ala
 165 170 175
 5 Leu Gly Ala Ser Tyr Ile Ala Leu Pro Trp Trp Ala Gly His Ala Leu
 180 185 190
 Phe Gly Thr Leu Asn Pro Thr Ile Met Val Leu Thr Leu Ile Tyr Ser
 195 200 205
 Leu Ala Gly Leu Gly Ile Ala Val Val Asn Asp Phe Lys Ser Val Glu
 210 215 220
 10 Gly Asp Arg Gln Leu Gly Leu Lys Ser Leu Pro Val Met Phe Gly Ile
 225 230 235 240
 Gly Thr Ala Ala Trp Ile Cys Val Ile Met Ile Asp Val Phe Gln Ala
 245 250 255
 15 Gly Ile Ala Gly Tyr Leu Ile Tyr Val His Gln Gln Leu Tyr Ala Thr
 260 265 270
 Ile Val Leu Leu Leu Leu Ile Pro Gln Ile Thr Phe Gln Asp Met Tyr
 275 280 285
 Phe Leu Arg Asn Pro Leu Glu Asn Asp Val Lys Tyr Gln Ala Ser Ala
 290 295 300
 20 Gln Pro Phe Leu Val Phe Gly Met Leu Ala Thr Gly Leu Ala Leu Gly
 305 310 315 320
 His Ala Gly Ile

 25
 <210> 35
 <211> 307
 <212> PRT
 <213> Synechocystis sp
 30
 <400> 35
 Met Thr Glu Ser Ser Pro Leu Ala Pro Ser Thr Ala Pro Ala Thr Arg
 1 5 10 15
 Lys Leu Trp Leu Ala Ala Ile Lys Pro Pro Met Tyr Thr Val Ala Val
 20 25 30
 35 Val Pro Ile Thr Val Gly Ser Ala Val Ala Tyr Gly Leu Thr Gly Gln
 35 40 45
 Trp His Gly Asp Val Phe Thr Ile Phe Leu Leu Ser Ala Ile Ala Ile
 50 55 60
 40 Ile Ala Trp Ile Asn Leu Ser Asn Asp Val Phe Asp Ser Asp Thr Gly
 65 70 75 80
 Ile Asp Val Arg Lys Ala His Ser Val Val Asn Leu Thr Gly Asn Arg
 85 90 95
 Asn Leu Val Phe Leu Ile Ser Asn Phe Phe Leu Leu Ala Gly Val Leu
 100 105 110
 45 Gly Leu Met Ser Met Ser Trp Arg Ala Gln Asp Trp Thr Val Leu Glu

115 120 125
 Leu Ile Gly Val Ala Ile Phe Leu Gly Tyr Thr Tyr Gln Gly Pro Pro
 130 135 140
 Phe Arg Leu Gly Tyr Leu Gly Leu Gly Glu Leu Ile Cys Leu Ile Thr
 5 145 150 155 160
 Phe Gly Pro Leu Ala Ile Ala Ala Tyr Tyr Ser Gln Ser Gln Ser
 165 170 175
 Phe Ser Trp Asn Leu Leu Thr Pro Ser Val Phe Val Gly Ile Ser Thr
 180 185 190
 10 Ala Ile Ile Leu Phe Cys Ser His Phe His Gln Val Glu Asp Asp Leu
 195 200 205
 Ala Ala Gly Lys Lys Ser Pro Ile Val Arg Leu Gly Thr Lys Leu Gly
 210 215 220
 Ser Gln Val Leu Thr Leu Ser Val Val Ser Leu Tyr Leu Ile Thr Ala
 15 225 230 235 240
 Ile Gly Val Leu Cys His Gln Ala Pro Trp Gln Thr Leu Leu Ile Ile
 245 250 255
 Ala Ser Leu Pro Trp Ala Val Gln Leu Ile Arg His Val Gly Gln Tyr
 260 265 270
 20 His Asp Gln Pro Glu Gln Val Ser Asn Cys Lys Phe Ile Ala Val Asn
 275 280 285
 Leu His Phe Phe Ser Gly Met Leu Met Ala Ala Gly Tyr Gly Trp Ala
 290 295 300
 Gly Leu Gly
 25 305

 <210> 36
 <211> 927
 <212> DNA
 30 <213> Synechocystis sp

 <400> 36
 atggcaacta tccaagcttt ttggcgcttc tcccgccccc ataccatcat tgggtacaact 60
 ctgagcgtct gggctgtgta tctgttaact attctcgggg atggaaactc agttaactcc 120
 35 cctgcttccc tggatttagt gttcggcgct tggctggcct gcctgttggg taatgtgtac 180
 attgtcggcc tcaaccaatt gtgggatgtg gacattgacc gcatcaataa gccgaatttg 240
 cccttagcta acggagattt ttctatcgcc cagggccggt ggattgtggg actttgtggc 300
 gttgcttcct tggcgatcgc ctggggatta gggctatggc tggggctaac ggtgggcatt 360
 agtttgatta ttggcacggc ctattcgggt cgcagtgga ggtaaagcg cttttccctg 420
 40 ctggcggccc tgtgtattct gacgggtcgg ggaattgtgg ttaacttggg cttattttta 480
 ttttttagaa ttggtttagg ttatccccc actttaataa ccccatctg ggttttgact 540
 ttatttatct tagttttcac cgtggcgatc gccattttta aagatgtgcc agatatggaa 600
 ggcgatcggc aatttaagat tcaaacttta actttgcaaa tcggcaaaaca aaacgttttt 660
 cggggaacct taattttact cactggttgt tatttagcca tggcaatctg gggcttatgg 720
 45 gcggctatgc ctttaaatac tgctttcttg attgtttccc atttgtgctt attagcctta 780
 ctctggtggc ggagtcgaga tgtacactta gaaagcaaaa ccgaaattgc tagtttttat 840

cagtttattt ggaagctatt tttcttagag tacttgctgt atcccttggc tctgtggta 900
cctaattttt ctaatactat tttttag 927

<210> 37

5 <211> 308

<212> PRT

<213> Synechocystis sp

<400> 37

10 Met Ala Thr Ile Gln Ala Phe Trp Arg Phe Ser Arg Pro His Thr Ile
1 5 10 15
Ile Gly Thr Thr Leu Ser Val Trp Ala Val Tyr Leu Leu Thr Ile Leu
20 25 30
Gly Asp Gly Asn Ser Val Asn Ser Pro Ala Ser Leu Asp Leu Val Phe
15 35 40 45
Gly Ala Trp Leu Ala Cys Leu Leu Gly Asn Val Tyr Ile Val Gly Leu
50 55 60
Asn Gln Leu Trp Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asn Leu
65 70 75 80
20 Pro Leu Ala Asn Gly Asp Phe Ser Ile Ala Gln Gly Arg Trp Ile Val
85 90 95
Gly Leu Cys Gly Val Ala Ser Leu Ala Ile Ala Trp Gly Leu Gly Leu
100 105 110
Trp Leu Gly Leu Thr Val Gly Ile Ser Leu Ile Ile Gly Thr Ala Tyr
25 115 120 125
Ser Val Pro Pro Val Arg Leu Lys Arg Phe Ser Leu Leu Ala Ala Leu
130 135 140
Cys Ile Leu Thr Val Arg Gly Ile Val Val Asn Leu Gly Leu Phe Leu
145 150 155 160
30 Phe Phe Arg Ile Gly Leu Gly Tyr Pro Pro Thr Leu Ile Thr Pro Ile
165 170 175
Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr Val Ala Ile Ala Ile
180 185 190
Phe Lys Asp Val Pro Asp Met Glu Gly Asp Arg Gln Phe Lys Ile Gln
35 195 200 205
Thr Leu Thr Leu Gln Ile Gly Lys Gln Asn Val Phe Arg Gly Thr Leu
210 215 220
Ile Leu Leu Thr Gly Cys Tyr Leu Ala Met Ala Ile Trp Gly Leu Trp
225 230 235 240
40 Ala Ala Met Pro Leu Asn Thr Ala Phe Leu Ile Val Ser His Leu Cys
245 250 255
Leu Leu Ala Leu Leu Trp Trp Arg Ser Arg Asp Val His Leu Glu Ser
260 265 270
Lys Thr Glu Ile Ala Ser Phe Tyr Gln Phe Ile Trp Lys Leu Phe Phe
45 275 280 285
Leu Glu Tyr Leu Leu Tyr Pro Leu Ala Leu Trp Leu Pro Asn Phe Ser

	290	295	300	
	Asn Thr Ile Phe			
	305			
5	<210> 38			
	<211> 1092			
	<212> DNA			
	<213> Synechocystis sp			
10	<400> 38			
	atgaaatttc cgccccacag tggttaccat tggcaaggtc aatcaccttt ctttgaaggt	60		
	tggtagctgc gcctgctttt gccccaatcc ggggaaagtt ttgcttttat gtactccatc	120		
	gaaaatcctg ctagcgatca tcattacggc ggcggtgctg tgcaaatctt agggccggt	180		
	acgaaaaaac aagaaaatca ggaagaccaa cttgtttggc ggacatttcc ctcggtaaaa	240		
15	aaattttggg ccagtcctcg ccagtttgcc ctagggcatt. ggggaaaatg tagggataac	300		
	aggcaggcga aaccctact ctccgaagaa ttttttgcca cggtaagga aggttatcaa	360		
	atccatcaaa atcagcacca aggacaaatc attcatggcg atcgccattg tcgttggcag	420		
	ttcaccgtag aaccggaagt aacttggggg agtcctaacc gatttcctcg ggctacagcg	480		
	ggttggcttt cctttttacc cttgtttgat cccggttggc aaattctttt agccaaggt	540		
20	agagcgacg gctggctgaa atggcagagg gaacagtatg aatttgacca cgccctagtt	600		
	tatgccgaaa aaaattgggg tcactccttt ccctcccgct ggttttggct ccaagcaaat	660		
	tattttcctg accatccagg actgagcgtc actgccgctg gcggggaacg gattgttctt	720		
	ggtcgccccg aagaggtagc ttaattggc ttacatcacc aaggtaattt ttacgaattt	780		
	ggcccgggcc atggcacagt cacttggcaa gtagctccct ggggcccgtt gcaattaaaa	840		
25	gccagcaatg ataggtattg ggtaagttg tccgaaaaaa cagataaaaa aggcagttta	900		
	gtccacactc ccaccgcca gggcttacia ctcaactgcc gagataccac taggggctat	960		
	ttgtatttgc aattgggac tgtgggtcac ggcctgatag tgcaagggga aacggacacc	1020		
	gcggggctag aagttggagg tgattgggg ttaacagagg aaaatttgag caaaaaaaca	1080		
	gtgccattct ga	1092		
30	<210> 39			
	<211> 363			
	<212> PRT			
	<213> Synechocystis sp			
35	<400> 39			
	Met Lys Phe Pro Pro His Ser Gly Tyr His Trp Gln Gly Gln Ser Pro			
	1 5 10 15			
	Phe Phe Glu Gly Trp Tyr Val Arg Leu Leu Leu Pro Gln Ser Gly Glu			
40	20 25 30			
	Ser Phe Ala Phe Met Tyr Ser Ile Glu Asn Pro Ala Ser Asp His His			
	35 40 45			
	Tyr Gly Gly Gly Ala Val Gln Ile Leu Gly Pro Ala Thr Lys Lys Gln			
	50 55 60			
45	Glu Asn Gln Glu Asp Gln Leu Val Trp Arg Thr Phe Pro Ser Val Lys			
	65 70 75 80			

Lys Phe Trp Ala Ser Pro Arg Gln Phe Ala Leu Gly His Trp Gly Lys
 85 90 95
 Cys Arg Asp Asn Arg Gln Ala Lys Pro Leu Leu Ser Glu Glu Phe Phe
 100 105 110
 5 Ala Thr Val Lys Glu Gly Tyr Gln Ile His Gln Asn Gln His Gln Gly
 115 120 125
 Gln Ile Ile His Gly Asp Arg His Cys Arg Trp Gln Phe Thr Val Glu
 130 135 140
 Pro Glu Val Thr Trp Gly Ser Pro Asn Arg Phe Pro Arg Ala Thr Ala
 10 145 150 155 160
 Gly Trp Leu Ser Phe Leu Pro Leu Phe Asp Pro Gly Trp Gln Ile Leu
 165 170 175
 Leu Ala Gln Gly Arg Ala His Gly Trp Leu Lys Trp Gln Arg Glu Gln
 180 185 190
 15 Tyr Glu Phe Asp His Ala Leu Val Tyr Ala Glu Lys Asn Trp Gly His
 195 200 205
 Ser Phe Pro Ser Arg Trp Phe Trp Leu Gln Ala Asn Tyr Phe Pro Asp
 210 215 220
 His Pro Gly Leu Ser Val Thr Ala Ala Gly Gly Glu Arg Ile Val Leu
 20 225 230 235 240
 Gly Arg Pro Glu Glu Val Ala Leu Ile Gly Leu His His Gln Gly Asn
 245 250 255
 Phe Tyr Glu Phe Gly Pro Gly His Gly Thr Val Thr Trp Gln Val Ala
 260 265 270
 25 Pro Trp Gly Arg Trp Gln Leu Lys Ala Ser Asn Asp Arg Tyr Trp Val
 275 280 285
 Lys Leu Ser Gly Lys Thr Asp Lys Lys Gly Ser Leu Val His Thr Pro
 290 295 300
 Thr Ala Gln Gly Leu Gln Leu Asn Cys Arg Asp Thr Thr Arg Gly Tyr
 30 305 310 315 320
 Leu Tyr Leu Gln Leu Gly Ser Val Gly His Gly Leu Ile Val Gln Gly
 325 330 335
 Glu Thr Asp Thr Ala Gly Leu Glu Val Gly Gly Asp Trp Gly Leu Thr
 340 345 350
 35 Glu Glu Asn Leu Ser Lys Lys Thr Val Pro Phe
 355 360

<210> 40

<211> 56

40 <212> DNA

<213> Artificial Sequence

<400> 40

45 cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat ttaaat

56

<210> 41

<211> 32
<212> DNA
<213> Artificial Sequence

5 <400> 41
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 42
<211> 32
10 <212> DNA
<213> Artificial Sequence

<400> 42
tcgacctgca ggaagcttgc ggccgcggat cc 32

15 <210> 43
<211> 32
<212> DNA
<213> Artificial Sequence

20 <400> 43
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 44
25 <211> 32
<212> DNA
<213> Artificial Sequence

<400> 44
30 tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 45
<211> 36
<212> DNA
35 <213> Artificial Sequence

<400> 45
tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

40 <210> 46
<211> 28
<212> DNA
<213> Artificial Sequence

45 <400> 46
cctgcaggaa gcttgccggcc gcggatcc 28

<210> 47
<211> 36
<212> DNA
5 <213> Artificial Sequence

<400> 47
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

10 <210> 48
<211> 28
<212> DNA
<213> Artificial Sequence

15 <400> 48
ggatccgcgg ccgcaagctt cctgcagg 28

<210> 49
<211> 39
20 <212> DNA
<213> Artificial Sequence

<400> 49
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

25 <210> 50
<211> 31
<212> DNA
<213> Artificial Sequence

30 <400> 50
ttggatccgc ggccgcaagc ttctgcagg t 31

<210> 51
35 <211> 41
<212> DNA
<213> Artificial Sequence

<400> 51
40 ggatccgcgg ccgcacaatg gagtctctgc tctctagttc t 41

<210> 52
<211> 38
<212> DNA
45 <213> Artificial Sequence

	<400> 52	
	ggatcctgca ggtcacttca aaaaaggtaa cagcaagt	38
5	<210> 53	
	<211> 45	
	<212> DNA	
	<213> Artificial Sequence	
10	<400> 53	
	ggatccgcgg ccgcacaatg gcgttttttg ggctctcccg tggtt	45
	<210> 54	
	<211> 40	
	<212> DNA	
15	<213> Artificial Sequence	
	<400> 54	
	ggatcctgca gggtattgaa aacttcttcc aagtacaact	40
20	<210> 55	
	<211> 38	
	<212> DNA	
	<213> Artificial Sequence	
25	<400> 55	
	ggatccgcgg ccgcacaatg tggcgaagat ctgttggt	38
	<210> 56	
	<211> 37	
30	<212> DNA	
	<213> Artificial Sequence	
	<400> 56	
35	ggatcctgca ggatcatggag agtagaagga aggagct	37
	<210> 57	
	<211> 50	
	<212> DNA	
	<213> Artificial Sequence	
40	<400> 57	
	ggatccgcgg ccgcacaatg gtacttgccg aggttcacaa gcttgccctt	50
	<210> 58	
45	<211> 38	
	<212> DNA	

<213> Artificial Sequence

<400> 58
5 ggatcctgca ggtcacttgt ttctggtgat gactctat 38

<210> 59
<211> 38
<212> DNA
<213> Artificial Sequence

10 <400> 59
ggatccgagg ccgcacaatg acttcgattc tcaacact 38

<210> 60
15 <211> 36
<212> DNA
<213> Artificial Sequence

<400> 60
20 ggatcctgca ggtcagtgtt gcatgctaa tgccgt 36

<210> 61
<211> 22
<212> DNA
25 <213> Artificial Sequence

<400> 61
taatgtgtac attgtcggcc tc 22

30 <210> 62
<211> 60
<212> DNA
<213> Artificial Sequence

35 <400> 62
gcaatgtaac atcagagatt ttgagacaca acgtggcttt ccacaattcc ccgcaccgtc 60

<210> 63
<211> 22
40 <212> DNA
<213> Artificial Sequence

<400> 63
aggctaataa gcacaaatgg ga 22

45 <210> 64

<211> 63
 <212> DNA
 <213> Artificial Sequence

5 <400> 64
 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggaattgg ttaggttat 60
 ccc 63

<210> 65
 10 <211> 26
 <212> DNA
 <213> Artificial Sequence

<400> 65
 15 ggatccatgg ttgcccaaac cccatc 26

<210> 66
 <211> 61
 <212> DNA
 20 <213> Artificial Sequence

<400> 66
 gcaatgtaac atcagagatt ttgagacaca acgtggcttt gggtaagcaa caatgaccgg 60
 c 61

25 <210> 67
 <211> 25
 <212> DNA
 <213> Artificial Sequence

30 <400> 67
 gaattctcaa agccagccca gtaac 25

<210> 68
 35 <211> 63
 <212> DNA
 <213> Artificial Sequence

<400> 68
 40 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcgggtgcga aaagggtttt 60
 ccc 63

<210> 69
 <211> 23
 45 <212> DNA
 <213> Artificial Sequence

1. <400> 69
ccagtgggttt aggctgtgtg gtc 23

5 <210> 70
<211> 21
<212> DNA
<213> Artifical Sequence

10 <400> 70
ctgagttgga tgtattggat c 21

<210> 71
<211> 28
15 <212> DNA
<213> Artifical Sequence

<400> 71
ggatccatgg ttacttcgac aaaaatcc 28

20 <210> 72
<211> 60
<212> DNA
<213> Artifical Sequence

25 <400> 72
gcaatgtaac atcagagatt ttgagacaca acgtggcttt gctaggcaac cgcttagtac 60

<210> 73
30 <211> 28
<212> DNA
<213> Artifical Sequence

<400> 73
35 gaattcttaa cccaacagta aagttccc 28

<210> 74
<211> 63
<212> DNA
40 <213> Artifical Sequence

<400> 74
ggtatgagtc agcaacacct tcttcacgag gcagacctca gcgcggcat tgtcttttac 60
atg 63

45 <210> 75

<211> 20
<212> DNA
<213> Artificial Sequence

5 <400> 75
ggaacccttg cagccgcttc 20

<210> 76
<211> 22
10 <212> DNA
<213> Artificial Sequence

<400> 76
gtatgcccaa ctggtgcaga gg 22

15 <210> 77
<211> 28
<212> DNA
<213> Artificial Sequence

20 <400> 77
ggatccatgt ctgacacaca aaataccg 28

<210> 78
25 <211> 62
<212> DNA
<213> Artificial Sequence

<400> 78
30 gcaatgtaac atcagagatt ttgagacaca acgtggcttt cgccaatacc agccaccaac 60
ag 62

<210> 79
<211> 27
35 <212> DNA
<213> Artificial Sequence

<400> 79
gaattctcaa atccccgcat ggcctag 27

40 <210> 80
<211> 65
<212> DNA
<213> Artificial Sequence

45 <400> 80

ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggcctacg gcttggacgt 60
gtggg 65

<210> 81
5 <211> 21
<212> DNA
<213> Artificial Sequence

<400> 81
10 cacttggatt cccctgatct g 21

<210> 82
<211> 21
<212> DNA
15 <213> Artificial Sequence

<400> 82
gcaatacccg cttggaaaac g 21

<210> 83
20 <211> 29
<212> DNA
<213> Artificial Sequence

<400> 83
25 ggatccatga ccgaatcttc gcccttagc 29

<210> 84
<211> 61
30 <212> DNA
<213> Artificial Sequence

<400> 84
35 gcaatgtaac atcagagatt ttgagacaca acgtggcttt caatcctagg tagccgaggc 60
g 61

<210> 85
<211> 27
<212> DNA
40 <213> Artificial Sequence

<400> 85
gaattcttag cccaggccag cccagcc 27

<210> 86
45 <211> 66

<212> DNA
<213> Artificial Sequence

<400> 86
5 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggggaatt gatttggtta 60
attacc 66

<210> 87
<211> 21
10 <212> DNA
<213> Artificial Sequence

<400> 87
15 gcgatcgcca ttatcgcttg g 21

<210> 88
<211> 24
<212> DNA
20 <213> Artificial Sequence

<400> 88
gcagactggc aattatcagt aacg 24

<210> 89
25 <211> 25
<212> DNA
<213> Artificial Sequence

<400> 89
30 ccatggattc gagtaaagtt gtcgc 25

<210> 90
<211> 0
<213> Artificial Sequence
35

<400> 90
gaattcactt caaaaaaggt aacag

<210> 91
40 <211> 4550
<212> DNA
<213> Arabidopsis sp

<400> 91
45 attttacacc aatttgatca cttaactaaa ttaattaaat tagatgatta tcccaccata 60
tttttgagca ttaaaccata aaaccatagt tataagtaac tgttttaatc gaatatgact 120

	cgattaagat	taggaaaaat	ttataaccgg	taattaagaa	aacattaacc	gtagtaaccg	180
	taaatgccga	ttcctccctt	gtctaaaaga	cagaaaacat	atattttatt	ttgccccata	240
	tgtttcactc	tatttaattt	caggcacaat	acttttggtt	ggtaacaaaa	ctaaaaagga	300
	caacacgtga	tacttttcct	cgtccgtcag	tcagattttt	tttaaaactag	aaacaagtgg	360
5	caaacttaca	ccacattttt	tgcttaactt	attaacttgt	aagtttttaa	ttcctaaaaa	420
	agtctaacta	attctttctaa	tataagtaca	ttccctaaat	ttcccaaaaa	gtcaaattaa	480
	taattttcaa	aatctaactt	aaatatctaa	taattcaaaa	tcattaaaaa	gacacgcaac	540
	aatgacacca	attaatcatt	ctcgacccac	acaattctac	agttctcatg	ctaaaccata	600
	ttttttgtct	tctgttcctt	caaaatcatt	tctttctctt	ctttgattcc	caaagatcac	660
10	ttctttgtct	ttgatttttg	attttttttt	tctctggcgt	gaaggaagaa	gctttatttc	720
	atggagtctc	tgctctctag	ttctctctct	gtttccgctg	gtaaatctcg	tccttttctg	780
	gtttcagggt	ttatttggtg	tttaggtttc	gtttttgtga	ttcagaacca	tacaaaaagt	840
	ttgaactttt	ctgaatataa	aataaggaaa	aagtttcgat	ttttataatg	aattgtttac	900
	tagatcgaag	taggtgacaa	agggtattgt	gtggagaagc	ataattttctg	ggcttgactt	960
15	tgaattttgt	ttctcatgca	tgcaacttat	caatcagctg	gtgggttttg	ttggaagaag	1020
	cagaatctaa	agctccactc	tttatcagggt	tcgttaggggt	tttatgggtt	tttgaaatta	1080
	aatactcaat	catcttagtc	tcattattct	attgggtgaa	tcacattttc	taatttgga	1140
	tttatgagac	aatgtatggt	ggacttagtt	gaagttcttc	tctttggtta	tagttgaagt	1200
	gttactgatg	ttgttttagct	ctttacacca	atatatacac	ccaattttgc	agaaatccga	1260
20	gttctgcgtt	gtgattcgag	ttaaagttgtc	gcaaaaccga	agtttaggaa	caatcttggt	1320
	aggcctgatg	gtcaaggatc	ttcattgttg	ttgtatccaa	aacataagtc	gagatttcgg	1380
	gttaaatgcc	ctgcgggtca	gcctgaggct	ttcgaactga	atagcaaaac	gaagtccttt	1440
	agagactcgt	tagatgcgtt	ttacagggtt	tctaggcctc	atacagttat	tggcacagtt	1500
	aagtttctct	ttaaaaatgt	aactctttta	aaacgcaatc	tttcagggtt	ttcaaggaga	1560
25	taacattagc	tctgtgattg	gatttgcagg	tgcttagcat	tttatctgta	tctttcttag	1620
	cagtagagaa	ggtttctgat	atatctcctt	tacttttcac	tggcatcttg	gaggtaatga	1680
	atatataaca	cataatgacc	gatgaagaag	atacattttt	ttcgtctctc	tgtttaaaca	1740
	attgggtttt	gttttcaggc	tggtgttgca	gctctcatga	tgaacattta	catagtggg	1800
	ctaaatcagt	tgtctgatgt	tgaatatagat	aaggtaacat	gcaaattttc	ttcatatgag	1860
30	ttcgagagac	tgatgagatt	aatagcagct	agtgccctaga	tcattctctat	gtgggttttt	1920
	gcagggttaac	aagccctatc	ttccattggc	atcaggagaa	tattctgtta	acaccggcat	1980
	tgcaatagta	gcttccttct	ccatcatggt	atggtgccat	tttcacaaaa	tttcaacttt	2040
	tagaattcta	taagttaactg	aaatagtttg	ttataaatcg	ttatagagtt	tctggcttgg	2100
	gtggattgtt	ggttcatggc	cattgttctg	ggctcttttt	gtgagtttca	tgctcggtac	2160
35	tgcatactct	atcaatgtaa	gtaagtttct	caatactaga	atttggtctca	aatcaaaatc	2220
	tgcaagttct	agtttttaggt	taatgaggtt	ttaataactt	acttctacta	caaacagttg	2280
	ccactttttac	ggtaggaaaag	atttgcattg	gttgcagcaa	tgtgtatcct	cgctgtccga	2340
	gctattattg	ttcaaatcgc	cttttatcta	catattcagg	tactaaacca	ttttccttat	2400
	gtttttagtg	tgttttcac	aaaatcactt	ttatattact	aaagctgtga	aactttgttg	2460
40	cagacacatg	tggtttggaag	accaatcttg	ttcactaggc	ctcttatttt	cgccactgcg	2520
	tttatgagct	ttttctctgt	cgttattgca	ttgtttaagg	taaacaaaga	tggaaaaaga	2580
	ttaaatctat	gtatacttaa	agtaaagcat	tctactgtta	ttgatgagaa	gttttctttt	2640
	ttggttggat	gcaggatata	cctgatatac	aaggggataa	gatattcggg	atccgatcat	2700
	tctctgtaac	tctgggtcag	aaacgggtac	gatattctaaa	ctaaagaaat	tgttttgact	2760
45	caagtgttgg	attaagatta	cagaagaaag	aaaactgttt	ttgtttcttg	caaaattcag	2820
	gtgttttgg	catgtgttac	actacttcaa	atggcttacg	ctgttgcaat	tctagtggga	2880

	gccacatctc cattcatatg gagcaaagtc atctcggtaa caatctttct ttacccatcg	2940
	aaaactcgct aattcatcgt ttgagtggta ctggtttcat tttgttccgt tctgttgatt	3000
	ttttttcagg ttgtgggtca tgttatactc gcaacaactt tgtgggctcg agctaagtc	3060
	gttgatctga gtagcaaaac cgaaataact tcattgttata tgttcatatg gaagggttaga	3120
5	ttcgtttata aatagagtct ttactgcctt tttatgctgt ccaatttggg attaaaatag	3180
	cctttcagtt tcatcgaatc accattatac tgataaattc tcatttctgc atcagctctt	3240
	ttatgcagag tacttgctgt tacctttttt gaagtgactg acattagaag agaagaagat	3300
	ggagataaaa gaataagtca tcactatgct tctgttttta ttacaagttc atgaaattag	3360
	gtagtgaact agtgaattag agttttatct tgaacatgg cagactgcaa aaatatgtca	3420
10	aagatatgaa tttctgttgg gtaaagaagt ctctgcttgg gcaaaatctt aagggttcgg	3480
	gtgttgatat aatgctaagc gaagaaatcg attctatgta gaaatttccg aaactatgtg	3540
	taaacatgtc agaacatctc cattctatat ctcttctgc aagaaagctc tgtttttatc	3600
	acctaaactc tttatctctg tgtagttaag atatgtatat gtacgtgact acattttttt	3660
	gttgatgtaa tttgcagaac gtatggattt ttgttagaaa gcatgagttc gaaagtatat	3720
15	gtttatatat atggataatt cagacctaac gtcgaagctc acaagcataa attcactact	3780
	atagtttctg ctgtaataga tagttccatt gatgtcttga aactgtacgt aactgcctgg	3840
	gcgttttctg gttgatactg actactgagt gttctttctg agtggtgtaa gtatacaaga	3900
	agaagaatat aggtctcacg gaacgactgt ggtggaagat gaaatggaga tcatcacgta	3960
	gcggctttgc caaagaccga gtcacgatcg agtctatgaa gtctttacag ctgctgatta	4020
20	tgattgacca ttgcttagag acgcattgga atcttactag ggacttgctt gggagtttct	4080
	tcaagtacgt gtcagatcat acgatgtagg agatttcacg gctttgatgt gtttgtttgg	4140
	agtcacaatg cttaatgggc ttattggccc aataatagct agctcttttg cttagccgt	4200
	ttcgtttctc ccctggtggg gagtattatt agggatgggt gtgaccaaag tcaccagacc	4260
	tagagtgaat ctagttagt cctagaccat ggcccatggc ttttatttgt aatttgaaaa	4320
25	atgaacaatt ctttttgtaa ggaaaacttt tatatagtag acgtttacta tatagaaact	4380
	agttgaacta acttcgtgca attgcataat aatgggtgtga aatagagggg gcaaaactca	4440
	ataaacattt cgacgtacca agagttcgaa acaataagca aaatagattt ttttgcttca	4500
	gactaatttg tacaatgaat ggtaataaaa ccattgaagc ttttattaat	4550
30	<210> 92	
	<211> 4450	
	<212> DNA	
	<213> Arabidopsis sp	
35	<400> 92	
	tttaggttac aaaatcaatg atattgcgta tgtcaactat aaaagccaaa agtaaagcct	60
	cttgtttgac cagaagggtca tgatcattgt atacatacag ccaaactacc tcctggaaga	120
	aaagacatgg atcccaaca acaacaatag ctctctttac aagaaccagt agtaactagt	180
	cactaatcta aaagagttaa gtttcagctt ttctggcaat ggctccttga tcatttcaat	240
40	cctgaaggag acccactttg tagcaagacc atgtcctctg tttcacttac agtgtgtctc	300
	aaaagtctac ttcaattctt catatatagg ttcttcacac tacagcttca tcctcattcg	360
	ttgacagaga gagagtcttt attgaaaact tcttccaagt acaactccac taaatataat	420
	agcaccaaac cacttggtcg acacaaatct gtacagatat aaaaacacta ttaggttttc	480
	caaggcaaat cacataattg gattgtgaaa gagtacaaaa gataaaccac aattttcata	540
45	ctttctactg cagtcagcac cagatgataa gtcagctgtc cctatttgcc atcctaactg	600
	tcctgatgca gcggccagtg atgcgtaata ttgccaccct taatcattag agcgagaaac	660

	aaaaagaatc	aaaagacagt	aatggaatt	aggaatcaca	aatgagtcct	tgtaaagttt	720
	attgagtacc	gagatctgca	ctgaatccag	aaagtgcagg	aaaacctatg	gatgctgtgc	780
	caaatccagt	taaccaaagc	tttgtattat	caccgaatct	aagggtgtgt	gacttaacac	840
	caacttttac	atcatcttct	ttgtcctgga	gacacaatat	attagacatt	agtccatgga	900
5	aaaaaaatga	tttaacctag	aatatctcaa	aattacttgc	ataaaaactg	aacttgagct	960
	gaaatttttg	gttcgtagct	tgtggcatat	actatttcat	tttcaatggg	ccacaaaggt	1020
	aactttcttt	tctcacttct	gttgcaaacg	ggaagacttt	tatggggcta	actcttcact	1080
	taaagtatat	aaatcagatg	gaaaagggtg	gagatcaggg	taattttctt	ctttatgatt	1140
	gacaaaagtc	gaacatcgaa	atggatgcat	ttgcatgaga	catgaaacaa	aagctgaaaa	1200
10	agaaatctgt	ggtggtgaag	ctagaaaaag	aaaacaaagc	aagcaatatg	cacacattga	1260
	gattaactac	tttgctactg	gtcataatca	aatagatttt	gaagctaaaa	aataaaaagt	1320
	gaatatacct	gatgtgcata	aatagtatca	taaacaaggg	tccagcagac	tccggagaga	1380
	tagagagggg	gtacaataga	tgggtctatg	cttcctttaa	ctgcagtcca	tcctaacaat	1440
	gctccccagt	ttatggtcaa	acctaaaaag	gcttgaggct	gcaattataa	aaacgaatca	1500
15	atcataagaa	aatcagaaaa	tatataatgt	ctaactttga	gaagccagaa	tagatttaaa	1560
	ttacccaaaa	tgtaaacctc	ttcataagtg	ggtaggaaaa	gacaagtaac	aaagatgaag	1620
	cccctaaaa	acggctgcag	aatatacata	ctgaaatgag	ctcaagtaga	aaagaatttg	1680
	atcacaaaa	taaagacaag	acctgagaac	atatcttcag	aatttggggc	aactacataa	1740
	gggtgaacca	tatgtgtatg	tgaattttta	aacaaacact	tgcaaatatg	cgactttagg	1800
20	gcaagtaaaa	aatccaaaca	aacctgtaat	tgtaagtgtg	gagaagaatc	cctaagccta	1860
	aaagcaactg	cagcccagag	aatccaatcc	cttgaaatgg	tgtcaaaaga	ccactggcga	1920
	taggtcttag	ttttgtacga	tcaacctgga	tataaaagaa	atttgtgaag	caacataatc	1980
	taaaacaaaa	caaccataca	aaatcttgag	ctttacatac	aagcaaccga	tctttgttta	2040
	tggagaatg	aatccagtta	catgaatgct	gtgtatctac	cctaactact	aaacacatat	2100
25	ttcaatcgaa	aaacatattc	caccttcacc	atatctaaca	cctgaagtct	ttcacttttt	2160
	gaacgaagtc	atcagaacat	gcagataagc	tattacccaa	aacagagata	tgactggaaa	2220
	tgttgtcgtg	aattgatcca	acatagaaaa	atcaagacca	gttcagatg	tcaaagcaat	2280
	aacactttcc	caccatggtt	acagaaacca	tagttacaca	aaacatgttt	cctaaaccaa	2340
	catactaaag	ggatatataa	atttgacatc	actttatcac	cataccataa	gatagcttaa	2400
30	aaacaaactg	acctttgtat	ctatgtcctg	atcaagcaga	tcatttatag	tacaaccagc	2460
	acctctaaga	agtaatgttc	cgcaacccaa	taaagccata	tatttaaaac	ttggaaggct	2520
	tccaggatca	gcagccaacg	caatcgacct	atacaacaat	gatggagatt	cagagtatcg	2580
	atctattttac	atagctcttg	aactagatcc	atgacgaaac	atggaaatc	gttataatat	2640
	ctaaagactt	ccaaacagat	tcctgagtaa	gaaacccagt	ggaactatag	tactgtaaca	2700
35	tatataaaat	caaagaaaac	tcagggtttat	agcattatcc	aatcctgatt	tctgccaatc	2760
	cttaaccact	ctcccatgct	atcaaaaacc	tcagctcaag	atcatactac	ctaattgcct	2820
	atgagctctt	gggaagatca	ttatggattt	gataactgaa	aaaagtaaca	gagaaatagc	2880
	agactgcaag	aactactcca	aacttctcca	ctgatatgta	tgtagtctaa	caataataaa	2940
	cagacataaa	ttctttttatc	aagcttcaag	agcaagttag	tcagaaaaca	tcacagccaa	3000
40	accaaccagg	aaaacacata	actttatcac	ataaaactaa	atttaatgta	atctgactta	3060
	acataaacca	tcctttggga	cgaaaggaaa	ctatataaac	atgcagtctt	tctttccctc	3120
	agctattctt	tcggatggat	tataatgaat	ctcaaaagtg	aaatgtcttg	attctcagct	3180
	acattactca	aaggcgaaga	taaacttacc	acatacaagg	ccacgcaagc	aaccaagttc	3240
	caatgggttt	atccaatcga	gcaagcttag	cataacctct	aacttcttct	ggtaaataca	3300
45	aatctatcca	agaagcttcc	ttaacaacaa	caccatcact	cttctcctta	tcacttttct	3360
	tcggctttcc	ctccaaaacc	gaagaagacg	acgacattcc	acaaattaat	ctgtaattcc	3420

	aaccaacacc	aaaaaacttc	tccatgatgca	attctcttcc	tttactccat	acttggtaat	3480
	tatcattcca	tgaaggataa	cacttagtga	aaggatttgt	gtaatgggta	gtcacaggat	3540
	tggaacagga	tttatgttgt	gattgcaaaa	gagcagagga	agaagatgga	gttacggaga	3600
	cggaagattt	caacaaccgt	cttgaaacac	gggagagccc	aaaaaacgcc	atctttgaga	3660
5	gaaattgttg	cctggaagaa	acaaagactt	gagatttcaa	acgtaagtga	attcttacga	3720
	acgaaagcta	acttctcaag	agaatcagat	tagtgattcc	tcaaaaacaa	acaaaactat	3780
	ctaatttcag	tttcgagtga	tgaagcctta	agaatctaga	acctccatgg	cgtttcta	3840
	ctctcagaga	taatcgaatt	ccttaaacaa	tcaaagctta	gaaagagaag	aacaacaaca	3900
	acaacaaaaa	aaatcagatt	aacaaccgac	cagagagcaa	cgacgacgcc	ggcgagaaa	3960
10	agcacgtcgt	ctcggagcaa	gacttcttct	ccagtaaccc	ggatggatcg	ttaatgggcc	4020
	tgtagattat	tatatgtggg	ccgaaacaat	tgggtcagca	aaaacttggg	ggataatgaa	4080
	gaaacacgta	cagtatgcat	ttaggtccca	aattaattgg	ccatataatt	cgaatcagat	4140
	aaactaatca	acccctacct	tacttatttc	tcactgtttt	tatttctacc	ttagtagttg	4200
	aagaaacact	tttatttatc	ttttcgggac	ccaaatttga	taggatcggg	ccattactca	4260
15	tgagcgtcag	acacatatta	gccttatcag	attagtgggg	taaggttttt	ttaattcggt	4320
	aagaagcaac	aatcaatgtc	ggagaaatta	aagaatctgc	atgggcgtgg	cgtgatgata	4380
	tgtgcatatg	gagtcagttg	ccgatcatat	ataactattt	ataaactaca	tataaagact	4440
	actaatagat						4450
20	<210>	93					
	<211>	2850					
	<212>	DNA					
	<213>	Arabidopsis sp					
25	<400>	93					
	aattaaaaatt	tgagcgggtct	aaaccattag	accgttttaga	gatccctcca	acccaaaata	60
	gtcgatttttc	acgtcttgaa	catatattgg	gccttaattct	gtgtgggttag	taaagacttt	120
	tatttggtcaa	agaaaaacaa	ccatggccca	acatgttgat	acttttattt	aattatacaa	180
	gtacccttga	attctctgaa	atataattga	ttgaccaga	tattaatttt	aattatcatt	240
30	tccctgtaaaa	gtgaaggagt	caccgtgact	cgctcgtaac	tgaaccaat	ctgttcatat	300
	gatgaagaag	tttctctcgt	tctctcccaa	cgcgtagaaa	attctgacgg	cttaacgatg	360
	tggcgaagat	ctgttggtta	tcgtttctct	tcaagaatct	ctgtttcttc	ttcgttacca	420
	aaccctagac	tgattccttg	gtcccgcgaa	ttatgtgccg	ttaatagctt	ctcccagcct	480
	ccggtctcga	cggaatcaac	tgctaagtta	gggatcactg	gtgttagatc	tgatgccaat	540
35	cgagtttttg	ccactgctac	tgccgcccgt	acagctacag	ctaccaccgg	tgagatttcg	600
	tctagagttg	cggctttggc	tggattaggg	catcactacg	ctcgttggtta	ttgggagctt	660
	tctaaagcta	aacttaggta	tgtgtttact	tttcttttct	catgaaaaat	ctgaaaattt	720
	ccaattgttg	gattcttaaa	ttctcatttg	ttttatgggt	gtagtatgct	tgtgggttga	780
	acttctggaa	ctgggtatat	tctgggtacg	ggaaatgctg	caattagctt	cccggggctt	840
40	tgttacacat	gtgcaggaac	catgatgatt	gctgcatctg	ctaattcctt	gaatcagggtc	900
	attgaaatgt	tgagaagttc	ataaatttcg	aatccttggt	gtgtttatgt	agttgatctt	960
	gcttgcttat	gtttatgtag	ttgaaaagtt	taaaaatttc	taatccttgg	tagttgatct	1020
	cgcttggttg	ttttttcatt	ttctagattt	ttgagataag	caatgattct	aagatgaaaa	1080
	gaacgatgct	aaggccattg	ccttcaggac	gtattagtgt	tccacacgct	gttgcagggg	1140
45	ctactattgc	tgggtgcttct	ggtgcttggt	tgttgccag	caagggtgaat	gtttgttttt	1200
	ttatatgtga	tttctttggt	ttatgaatgg	gtgattgaga	gattatggat	ctaaactttt	1260

	gcttccacga caaggttatt gcagactaat atgttggctg ctggacttgc atctgccaat	1320
	cttgactttt atgcgtttgt ttatactccg ttgaagcaac ttcaccctat caatacatgg	1380
	gttggcgctg ttgttgggtg tatcccaccc ttgcttgggt aaatttttgt tccttttctt	1440
	ctttatttta gcagattctg ttttgttggg tactgctttt aattcaaaat gtagtcatgg	1500
5	ttcaccaatt ctatgcttat ctattttgtg tgttgtcagg tgggcggcag cgtctgggtca	1560
	gatttcatac aattcgatga ttcttccagc tgctctttac ttttggcaga tacctcattt	1620
	tatggccctt gcacatctct gccgcaatga ttatgcagct ggagggttaag accatatggg	1680
	gtcatatgag attagaatgt ctcttcccat gtagtgttga tcttgaacta gttcaatttc	1740
	gtggaatgat cagagtgtcc tagatagtgt cacagcagtc gacatttttag tggctagata	1800
10	atgagttctt tccgttagag ataaacattc gcgaacattg tttccagctt ccgcgaccca	1860
	acttctgatt ttgtttcttg gtaccttgtt ttcagttaca agatgttgtc actctttgat	1920
	ccgtcaggga agagaatagc agcagtgggt ctaagggaact gcttttacct gatccctctc	1980
	ggtttcatcg cctatgactg tgagtcttgt agattcatct tttttttgta gtttattgac	2040
	tgcattgctg tatctgattt ttgctgttcc ttccaatttt tgtgacaggg ggggttaacct	2100
15	caagttgggt ttgcctcgaa tcaacacttc tcacactagc aatcgctgca acagcatttt	2160
	cattctaccg agaccggacc atgcataaag caaggaaaat gttccatgcc agtcttctct	2220
	tccttctctg tttcatgtct ggtcttcttc tacaccgtgt ctctaattgat aatcagcaac	2280
	aactcgtaga agaagccgga ttaacaaatt ctgtatctgg tgaagtcaaa actcagaggc	2340
	gaaagaaacg tgtggctcaa cctccggtgg cttatgcctc tgctgcaccg tttcttttcc	2400
20	tcccagctcc ttccttctac tctccatgat aacctttaag caagctattg aatttttgga	2460
	aacagaaatt aaaaaaaaaa tctgaaaagt tcttaagttt aatctttggg taataatgaa	2520
	gtggagaacg catacaagtt tatgtatttt ttctcatctc cacataattg tattttttct	2580
	ctaagtatgt ttcaaattgat acaaaatata tactttatca attatctgat caaattgatg	2640
	aatttttgag ctttgacgtg ttaggtctat ctaataaacg tagtaacgaa tttgggtttg	2700
25	gaaatgaaat ccgataaccg atgatgggtg agagttaaac gattaaaccg gggtggttaa	2760
	aggtctcgag tctcgacggc tgcggaaatc ggaaaatcac gattgaggac tttgagctgc	2820
	cacgaagatg gcgatgaggt tgaaatcaat	2850
	<210> 94	
30	<211> 3660	
	<212> DNA	
	<213> Arabidopsis sp	
	<400> 94	
35	tattttgtatt tttattgtta aattttatga tttcaccggg tatatatcat cccatattaa	60
	tatttagattt attttttggg ctttatttgg gttttcgatt taaactgggc ccattctgct	120
	tcaatgaaac cctaattgggt tttgtttggg ctttggattt aaaccggggc cattctgctt	180
	caatgaagggt cttttgtcca acaaaactaa catccgacac aactagtatt gccaagagga	240
	tcgtgccaca tggcagttat tgaatcaaag gccgccaaaa ctgtaacgta gacattactt	300
40	atctccggta acggacaacc actcgtttcc cgaaacagca actcacagac tcacaccact	360
	ccagtctccg gcttaactac caccagagac gattctctct tccgtcgggt ctatgacttc	420
	gattctcaac actgtctcca ccattccact ttccagagtt acctccgtcg atcgagtcgg	480
	agtcctctct cttcggaatt cggattccgt tgagttcact cgcggcggtt ctgggtttctc	540
	gacgttgatc tacgaatcac ccggtagtta gcattctgtt ggatagattg atgaatgttt	600
45	tcttcgattt tttttttact gatcttgttg tggatctctc gtagggcgga gatttgttgt	660
	gcgtgcggcg gagactgata ctgataaagg tatgattttt tagttgtttt tattttctct	720

	ctcttcaaaa	ttctcttttc	aaacactgtg	gcgtttgaat	ttccgacggc	agttaaatct	780
	cagacacctg	acaaggcacc	agccggtggg	tcaagcatta	accagcttct	cggtatcaaa	840
	ggagcatctc	aagaaactgt	aattttgttc	atctcctcag	aatcttttaa	attatcatat	900
	ttgtggataa	tgatgtgtta	gtttaggaat	tttctacta	aaggtaatct	cttttgagga	960
5	caagtcttgt	ttttagctta	gaaatgatgt	gaaaatgttg	ttgttagct	aaaaagagtt	1020
	tgttgttata	ttctgtatc	agaataaatg	gaagattcgt	cttcagctta	caaaaccagt	1080
	cacttggcct	ccactgggtt	ggggagtcgt	ctgtggtgct	gctgcttcag	gtaatcatat	1140
	gaacctcttt	tggatcatgc	aatactgtac	agaaagtgtt	ttcattttcc	ttccaattgt	1200
	ttcttctggc	agggaaactt	cattggaccc	cagaggatgt	tgctaagtcg	attctttgca	1260
10	tgatgatgtc	tggtccttgt	cttactggct	atacacaggt	ctgggtttac	acaacaaaaa	1320
	gctgacttgt	tcttattcta	gtgcatttgc	ttggtgctac	aataacctag	acttgtcgat	1380
	ttccagacaa	tcaacgactg	gtatgataga	gatatcgacg	caattaatga	gccatatact	1440
	ccaattccat	ctggagcaat	atcagagcca	gaggtaactg	agacagaaca	ttgtgagctt	1500
	ttatctcttt	tgtgattctg	atttctcctt	actccttaa	atgcagggtta	ttacacaagt	1560
15	ctgggtgcta	ttattgggag	gtcttggtat	tgctggaata	ttagatgtgt	gggtaagttg	1620
	gcccttctga	cattaactag	tacagttaaa	gggcacatca	gatttgctaa	aatcttccct	1680
	tatcaggcag	ggcataccac	tcccactgtc	ttctatcttg	ctttgggagg	atcattgcta	1740
	tcttatatat	actctgtctc	acctcttaag	gtaagtttta	ttcctaactt	ccactctcta	1800
	gtgataagac	actccatcca	agttttggag	ttttgaatat	cgatatctga	actgatctca	1860
20	ttgcagctaa	aacaaaatgg	atgggttggg	aattttgcac	ttggagcaag	ctatattagt	1920
	ttgccatggt	aagatatctc	gtgtatcaat	aatatatggc	gttgttctca	tctcattgat	1980
	ttgtttcttg	ctcacttgac	tgataggtgg	gctggccaag	cattgttttg	cactcttacg	2040
	ccagatgttg	ttgttctaac	actcttgtac	agcatagctg	gggtactctt	ttggcaaac	2100
	ttttatgttg	cttttttctg	tatctgttgt	aatatgctct	tgcttcatgt	tgtacctttg	2160
25	tgataatgca	gttaggaata	gccattgtta	acgacttcaa	aagtgttgaa	ggagatagag	2220
	cattaggact	tcagtctctc	ccagtagctt	ttggcaccga	aactgcaaaa	tggtatagcg	2280
	ttggtgctat	agacattact	cagctttctg	ttgccggtat	gtactatcca	ctgtttttgt	2340
	gcagctgtgg	cttctatttc	tttctcctga	tcttatcaac	tggtatattca	ccaatggtaa	2400
	agacaaaatt	aatgaagctg	aatcaacaaa	ggcaaaacat	aaaagtacat	tctaattgaa	2460
30	tgagctaatt	aagaggaggc	atctactttt	atgtttcatt	agtgtgattg	atggattttc	2520
	atttcatgct	tctaaaacaa	gtattttcaa	cagtgtcatg	aaataacaga	acttatatct	2580
	tcatttgtac	ttttactagt	ggatgagtta	cacaatcatt	gttatagaac	caaatcaaa	2640
	gtagagatca	tcattagtat	atgtctatct	tggttgacag	atatctatta	gcattctggg	2700
	aaccttatta	tgcgttggcg	ttggttgctt	tgatcattcc	tcagattgtg	ttccaggtaa	2760
35	agacgttaac	agtctcacat	tataattaat	caaattcttg	tcactcgtct	gattgtctaca	2820
	ctcgtttcta	taaactgcag	tttaataact	ttctcaagga	ccctgtcaaa	tacgacgtca	2880
	agtaccaggt	aagtcaactt	agtacacatg	tttgtgttct	tttgaaatat	ctttgagagg	2940
	tctcttaatc	agaagtgtgt	tgaacacttc	atcttgatta	caggcaagcg	cgagccatt	3000
	cttgggtgctc	ggaatatatt	taacggcatt	agcatcgcaa	cactgaaaaa	ggcgattttt	3060
40	gatgggggtt	tgctgaaagc	agagggtgtg	acacatcaaa	tggtgggcaag	tgatggcatc	3120
	aactagttta	aaagattttg	taaaatgtat	gtaccgttat	tactagaaac	aactcctgtt	3180
	gtatcaattt	agcaaaacgg	ctgagaaatt	gtaattgatg	ttaccgtatt	tgcgctccat	3240
	ttttgcattt	cctgctcata	tcgaggattg	gggtttatgt	tagttctgtc	acttctctgc	3300
	tttcagaatg	tttttgtttt	ctgtagtggg	ttttaactat	tttcatcact	ttttgtattg	3360
45	attctaaaca	tgtatccaca	taaaaacagt	aatatacaaa	aatgatactt	cctcaaaact	3420
	tttataatct	aaatctaaca	actagctagt	aacccaacta	acttcataca	attaatttga	3480

gaaactacaa	agactagact	atacatatgt	tatttaacaa	cttgaaactg	tggtattact	3540
acctgatttt	tttctattct	acagccattt	gatatgctgc	aatcttaaca	tatcaagtct	3600
cacgttggtg	gacacaacat	actatcacia	gtaagacacg	aagtaaaacc	aaccggcaac	3660

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 October 2000 (26.10.2000)

PCT

(10) International Publication Number
WO 00/63391 A3

- (51) International Patent Classification⁷: C12N 15/54.
15/82, 9/10, 5/00, C12P 17/06
- (21) International Application Number: PCT/US00/10368
- (22) International Filing Date: 14 April 2000 (14.04.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/129,899 15 April 1999 (15.04.1999) US
60/146,461 30 July 1999 (30.07.1999) US
- (71) Applicant (for all designated States except US): CAL-
GENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616
(US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): SAVIDGE, Beth
[US/US]; 1920 Fifth Street, Davis, CA 95616 (US).
LASSNER, Michael, W. [US/US]; 1920 Fifth Street,
Davis, CA 95616 (US). WEISS, James, D. [US/US];
800 N. Lindbergh Blvd., St. Louis, MO 63167 (US).
- POST-BEITTENMILLER, Dusty [US/US]; 800 N.
Lindbergh Blvd., St. Louis, MO 63167 (US).
- (74) Agent: RAE-VENTER LAW GROUP, P.C.; P.O. Box
60039, Palo Alto, CA 94306 (US).
- (81) Designated States (national): AL, AM, AT, AU, AZ, BA,
BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES,
FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG,
MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE,
SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN,
YU, ZW.
- (84) Designated States (regional): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).
- Published:
— with international search report
- (88) Date of publication of the international search report:
17 January 2002
- For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 00/63391 A3

(54) Title: NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS

(57) Abstract: Nucleic acid sequences and methods are provided for producing plants and seeds having altered tocopherol content and compositions. The methods find particular use in increasing the tocopherol levels in plants, and in providing desirable tocopherol compositions in a host plant cell.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/10368

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/54 C12N15/82 C12N9/10 C12N5/00 C12P17/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, WPI Data, EPO-Internal, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] ACCESSION NO: AC003673, 11 December 1997 (1997-12-11) LIN, X., ET AL.: "Arabidopsis thaliana chromosome II section 110 of 255 of the complete sequence. Sequence from clones MSF3, F19F24." XP002153685 nts40740-43320	1-6, 13-16,18
X	DATABASE EMBL [Online] ACCESSION NO: AL035394, 9 February 1999 (1999-02-09) BEVAN, M., ET AL.: "Arabidopsis thaliana DNA chromosome 4, BAC clone F9D16 (ESSAII project)" XP002153686 nts 46219-49152	1-6, 13-16,18

-/--

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

12 June 2001

Date of mailing of the international search report

15.06.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Maddox, A

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online] ACCESSION NO: B24116, 13 October 1997 (1997-10-13) ROUNSLEY, S.D., ET AL.: "F18L14TF IGF Arabidopsis thaliana genomic clone F18L14, genomic survey sequence." XP002153687 the whole document</p> <p>---</p>	1-6
X	<p>DATABASE EMBL [Online] ACCESSION NO: AC003672, 11 December 1997 (1997-12-11) LIN, X., ET AL.: "Arabidopsis thaliana chromosome II section 239 of 255 of the complete sequence. Sequence from clones F411, F16B22." XP002153688 nts 363-2540</p> <p>---</p>	1-6, 13-16,18
X	<p>DATABASE EMBL [Online] ACCESSION NO: B29398, 13 October 1997 (1997-10-13) ROUNSLEY, S.D., ET AL.: "F16B22TRC IGF Arabidopsis thaliana genomic clone F16B22, genomic survey sequence." XP002153689 the whole document</p> <p>---</p>	1-6
X	<p>DATABASE EMBL [Online] ACCESSION NO: R30625, 11 August 1995 (1995-08-11) NEWMAN, T., ET AL.: "13230 Lambda-PRL2 Arabidopsis thaliana cDNA clone 166L10T7, mRNA sequence." XP002153690 abstract</p> <p>---</p>	1-6
X	<p>GAUBIER PASCALE ET AL: "A chlorophyll synthetase gene from Arabidopsis thaliana." MOLECULAR & GENERAL GENETICS, vol. 249, no. 1, 1995, pages 58-64, XP002153682 ISSN: 0026-8925 the whole document -& DATABASE TREMBL [Online] ACCESSION NO: Q38833, 1 November 1996 (1996-11-01) GAUBIER, P., ET AL.: "PUTATIVE CHLOROPHYLL SYNTHETASE (G4)." XP002169117 the whole document</p> <p>---</p> <p style="text-align: center;">-/--</p>	1-6, 13-16,18

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE BIOSIS [Online] BIOSCIENCES INFORMATION SERVICE, PHILADELPHIA, PA, US; October 1997 (1997-10) OSTER U ET AL: "The G4 gene of Arabidopsis thaliana encodes a chlorophyll synthase of etiolated plants." Database accession no. PREV199800047824 XP002153691 abstract & BOTANICA ACTA, vol. 110, no. 5, October 1997 (1997-10), pages 420-423, ISSN: 0932-8629</p>	1-5, 13-16
X	<p>LOPEZ J ET AL: "Sequence of the bchG gene from Chloroflexus aurantiacus: Relationship between Chlorophyll synthase and other polyprenyltransferases" JOURNAL OF BACTERIOLOGY, WASHINGTON, DC, US, vol. 178, no. 11, 1996, pages 3369-3373, XP002146399 ISSN: 0021-9193 the whole document</p>	1-4
X	<p>DATABASE EMBL [Online] ACCESSION NO: AC004077, 3 February 1998 (1998-02-03) LIN, X., ET AL.: "Arabidopsis thaliana chromosome II section 190 of 255 of the complete sequence. Sequence from clones F13P17, T31E10." XP002169118 /gene="At2g34630" -& DATABASE TREMBL [Online] ACCESSION NO: 064684, 1 August 1998 (1998-08-01) ROUNSLEY S.D., ET AL.: "T31E10.3 PROTEIN" XP002169119 abstract</p>	1-6, 13-16, 18
X	<p>DATABASE EMBL [Online] ACCESSION NO: T44803, 4 February 1995 (1995-02-04) NEWMAN, T. ET AL.: "8066 Lambda-PRL2 Arabidopsis thaliana cDNA clone 124L9T7, mRNA sequence." XP002169120 the whole document</p>	1-5

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online] ACCESSION NO: Z34566, 25 June 1994 (1994-06-25) DESPREZ, T., ET AL.: "A. thaliana transcribed sequence; clone VBVC03; 5' end; Similar to Cytochrome c554 ; Chloroflexus aurantiacus." XP002169121 the whole document</p> <p>---</p>	1-5
X	<p>ZHU XUFEN ET AL: "Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from Arabidopsis thaliana is localized in mitochondria." PLANT MOLECULAR BIOLOGY, vol. 35, no. 3, 1997, pages 331-341, XP002153683 ISSN: 0167-4412 the whole document</p> <p>---</p>	1-5, 13-16, 29,30
X	<p>DATABASE EMBL [Online] ACCESSION NO: L40577, 15 April 1995 (1995-04-15) SCOLNIK, P.A., ET AL.: "Arabidopsis thaliana geranylgeranyl pyrophosphate synthase-related protein mRNA, complete cds." XP002153692 the whole document</p> <p>---</p>	1-5
X	<p>CHUN P L ET AL: "Identification of a maize endosperm-specific cDNA encoding farnesyl pyrophosphate synthetase" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 171, no. 2, 1 June 1996 (1996-06-01), pages 193-196; XP004042793 ISSN: 0378-1119 the whole document</p> <p>---</p>	1-4,7
E	<p>EP 1 033 405 A (CERES INC) 6 September 2000 (2000-09-06) see SEQ ID NOS:34834-34836,38169 38171,50712,50713</p> <p>---</p>	1-7, 13-16,18
E	<p>WO 00 68393 A (PIONEER HI-BRED) 16 November 2000 (2000-11-16)</p> <p>see SEQ ID NOS: 1,2,3,4,9-14,21,22,23</p> <p>---</p> <p>-/--</p>	1-16, 18-20, 22-25, 27-30, 32,33

INTERNATIONAL SEARCH REPORT

Interr. Application No

PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 14207 A (DU PONT ;MIAO GUO HUA (US); POWELL WAYNE (US); CAHOON REBECCA E (U) 16 March 2000 (2000-03-16) SEQ ID NOS:1,2,5,6,7,8,11 and 12 ---	1-4,7,9
P,X	DATABASE EMBL [Online] ACCESSION NO: AI795655, 7 July 1999 (1999-07-07) WALBOT, V.: "614004H08.x4 614 - root cDNA library from Walbot Lab Zea mays cDNA, mRNA sequence." XP002169131 the whole document ---	1-4,7
P,X	DATABASE EMBL [Online] ACCESSION NO: AI988542, 7 September 1999 (1999-09-07) SHOEMAKER, R., ET AL.: "sd03g09.y1 Gm-cl020 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:Gm-cl020-665 5' similar to TR:064886 064886 PUTATIVE HEME A:FARNESYLTRANSFERASE. ;, mRNA sequence." XP002169132 the whole document ---	1-4,9
P,X	DATABASE EMBL [Online] ACCESSION NO: AI938569, 3 August 1999 (1999-08-03) SHOEMAKER, R., ET AL.: "sb55e11.y1 Gm-cl018 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:Gm-cl018-69 5' similar to TR:064625 064625 F19F24.15 PROTEIN. ;, mRNA sequence." XP002169133 the whole document ---	1-4,9
P,X	DATABASE EMBL [Online] ACCESSION NO: AW306617, 21 January 2000 (2000-01-21) SHOEMAKER R., ET AL.: "se53b09.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:Gm-cl017-2610 5' similar to TR:064625 064625 F19F24.15 PROTEIN. ;, mRNA sequence" XP002169134 the whole document --- -/--	1-4,9

INTERNATIONAL SEARCH REPORT

Interr. Application No
PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>DATABASE EMBL [Online] ACCESSION NO: AI748688, 29 June 1999 (1999-06-29) SHOEMAKER R.: "sb60f03.y1 Gm-cl010 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:Gm-cl010-150 5' similar to TR:P73726 P73726 HYPOTHETICAL 34.4 KD PROTEIN.;, mRNA sequence." XP002169135 the whole document</p>	1-4,9
X	<p>--- DATABAS EMBL [Online] ACCESSION NO: D64006, 30 September 1995 (1995-09-30) TABATA, S., ET AL.: "Synechocystis sp. PCC6803 complete genome, 25/27, 3138604-3270709." XP002169122 nts 90109-90987 -& DATABAS TREMBL [Online] ACCESSION NO: Q55500, 1 November 1996 (1996-11-01) TABATA, S., : "4-HYDROXYBENZOATE-OCTAPRENYL TRANSFERASE." XP002169123 the whole document</p>	1-4,11, 12
X	<p>--- DATABAS EMBL [Online] ACCESSION NO: D13960, 28 March 1996 (1996-03-28) MURATA N.; ET AL.: "Synechocystis sp. genes for heme O synthase and virginiamycin acetyltransferase, complete cds." XP002169124 the whole document -& DATABAS TREMBL [Online] ACCESSION NO: Q55207, 1 November 1996 (1996-11-01) "CYTOSHROME C OXIDASE FOLDING PROTEIN" XP002169125 abstract</p> <p>--- -/--</p>	1-4,11, 12

INTERNATIONAL SEARCH REPORT

 Intern. Application No
 PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] ACCESSION NO: D64001, 30 September 1995 (1995-09-30) TABATA, S., ET AL.: "Synechocystis sp. PCC6803 complete genome, 20/27, 2539000-2644794." XP002169126 nts 39188..40162 -& DATABASE TREMBL [Online] ACCESSION NO: Q55145, 1 November 1996 (1996-11-01) TABATA, S.: "CHLOROPHYLL SYNTHASE 33 KDA SUBUNIT." XP002169127 the whole document	1-4,11, 12
X	--- DATABASE EMBL [Online] ACCESSION NO: D90911, 31 October 1996 (1996-10-31) TABATA S.;: "Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643." XP002169128 nts 33234..34157 -& DATABASE TREMBL [Online] ACCESSION NO: P73962, 15 July 1998 (1998-07-15) KANEKO, T., ET AL.: "PROBABLE 1,4-DIHYDROXY-2-NAPHTHOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-)" XP002169129 abstract	1-4,11, 12
X	--- DATABASE EMBL [Online] ACCESSION NO: D90909, 31 October 1996 (1996-10-31) TABATA, S.: "Synechocystis sp. PCC6803 complete genome, 11/27, 1311235-1430418." XP002169130 nts 11453..12379 and 12438..13529 -& DATABASE TREMBL [Online] ACCESSION NO: P73726, 1 February 1997 (1997-02-01) KANEKO, T., ET AL.: "HYPOTHETICAL 34.4 KDA PROTEIN." XP002169263 abstract -& DATABASE TREMBL [Online] ACCESSION NO: P73727, 1 February 1997 (1997-02-01) KANEKO, T., ET AL.: "HYPOTHETICAL 41.5 KDA PROTEIN." XP002169264 abstract --- -/--	1-4,11, 12

5

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interr. Application No
PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 07867 A (CALGENE LLC) 18 February 1999 (1999-02-18) page 11, line 19 - line 26 ---	1-4, 13-15, 18-20, 22-25, 27-30, 32,33
X	WO 98 06862 A (SHEWMAKER CHRISTINE K ;CALGENE INC (US)) 19 February 1998 (1998-02-19) the whole document ---	1-4, 13-15, 18,29, 30,32,33 19,20, 22-25, 27,28
A		
E	WO 00 61771 A (MONSANTO CO) 19 October 2000 (2000-10-19) page 81 -page 83 page 106 -page 107 ---	1-4, 13-15, 18-20, 22-25, 27-30, 32,33
A	NORRIS S R ET AL: "GENETIC DISSECTION OF CAROTENOID SYNTHESIS IN ARABIDOPSIS DEFINES PLASTOQUINONE AS AN ESSENTIAL COMPONENT OF PHYTOENE DESATURATION" PLANT CELL,US,AMERICAN SOCIETY OF PLANT PHYSIOLOGISTS, ROCKVILLE, MD, vol. 7, 1 December 1995 (1995-12-01), pages 2139-2149, XP002041909 ISSN: 1040-4651 the whole document ---	1-6, 13-16, 18-20, 22-25, 27-30, 32,33
X	KUNTZ, M., ET AL.: "Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening" THE PLANT JOURNAL, vol. 2, no. 1, 1992, XP002153684 the whole document ---	1-4, 13-15, 18,29,30
X	US 5 876 964 A (GERSHENZON JONATHAN ET AL) 2 March 1999 (1999-03-02) the whole document ---	1-4, 13-15, 18,29,30
X	US 5 545 816 A (AUSICH RODNEY L ET AL) 13 August 1996 (1996-08-13) the whole document ---	1-4, 13-15, 18,29,30

6

-/--

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interr. Application No
PCT/US 00/10368

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 763 542 A (TOYOTA MOTOR CO LTD) 19 March 1997 (1997-03-19) the whole document ---	29,30
X	EP 0 674 000 A (TOYOTA MOTOR CO LTD) 27 September 1995 (1995-09-27) the whole document ---	29,30
P,X	WO 00 01650 A (DCV INC) 13 January 2000 (2000-01-13) the whole document ---	1-4, 13-15, 18,29,30
E	EP 1 063 297 A (KOREA KUMHO PETROCHEM CO LTD) 27 December 2000 (2000-12-27) the whole document ---	1-4, 13-15, 18,29,30
A	WO 97 27285 A (UNIV ARIZONA) 31 July 1997 (1997-07-31) the whole document ---	19,20, 22-25, 27-30, 32,33
A	WO 99 04622 A (UNIV NEVADA) 4 February 1999 (1999-02-04) the whole document ---	19,20, 22-25, 27-30, 32,33
X	WO 99 06580 A (BONETTA DARIO ;MCCOURT PETER (CA); GHASSEMIAN MAJID (CA); PERFORMA) 11 February 1999 (1999-02-11) claims 18,19 ---	1-5, 13-16,18
E	WO 00 22150 A (PIONEER HI BRED INT ;YALPANI NASSER (US); MEYER TERRY EUCLAIRE (US)) 20 April 2000 (2000-04-20) the whole document ---	1-4, 13-15,18
E	WO 01 21650 A (COLDREN CHRIS ;DU PONT (US); WANG HONG (US); FLINT DENNIS (US); HA) 29 March 2001 (2001-03-29) the whole document -----	1-4, 13-16,18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/10368

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
Claims 8 and 10 are inconsistent with figs 2,3, and 9, since said figures do not make reference to sequence data. Said claims have been assumed as relating to SEQ ID NOS:19-31 and the prior art search has been made accordingly.
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-6,13-16, and 18 all partially

Nucleic acid sequences encoding Arabidopsis aromatic prenyltransferases as defined by SEQ ID NOS:1-6 and constructs based on said sequences.

2. Claims: 1-6,13-16, and 18 all partially.

Nucleic acid sequences encoding Arabidopsis straight chain prenyltransferases as defined by SEQ ID NOS:11,12,16,17 and constructs based on said sequences.

3. Claims: 1-4,13-16 and 18 all partially,
and 7 and 8 both completely

Nucleic acid sequences encoding corn prenyltransferases and constructs based on said sequences.

4. Claims: 1-4,13-16, and 18 all partially,
and 9 and 10 both completely

Nucleic acid sequences encoding soybean prenyltransferases and constructs based on said sequences.

5. Claims: 1-4,13,14, and 18 all partially, and 11,12,
and 17 all completely

Nucleic acid sequences encoding synechocystis prenyltransferases and constructs based on said sequences.

6. Claims: 19-33 all completely

Methods for production of tocopherols, increasing flux to tocopherol production in a host cell by transforming with a prenyltransferase nucleic acid sequence

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/10368

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1033405 A	06-09-2000	NONE	
WO 0068393 A	16-11-2000	AU 4498500 A	21-11-2000
WO 0014207 A	16-03-2000	AU 5812199 A	27-03-2000
WO 9907867 A	18-02-1999	AU 8900298 A CN 1275166 T EP 1002117 A	01-03-1999 29-11-2000 24-05-2000
WO 9806862 A	19-02-1998	AU 4058497 A BR 9713462 A CN 1227609 A EP 0925366 A	06-03-1998 28-03-2000 01-09-1999 30-06-1999
WO 0061771 A	19-10-2000	AU 4231600 A	14-11-2000
US 5876964 A	02-03-1999	AU 1089099 A EP 1023436 A WO 0129188 A WO 9919460 A	03-05-1999 02-08-2000 26-04-2001 22-04-1999
US 5545816 A	13-08-1996	US 5618988 A CA 2055447 A EP 0471056 A JP 5504686 T WO 9113078 A US 5530188 A US 5530189 A US 5684238 A US 5656472 A	08-04-1997 03-09-1991 19-02-1992 22-07-1993 05-09-1991 25-06-1996 25-06-1996 04-11-1997 12-08-1997
EP 0763542 A	19-03-1997	JP 9065878 A DE 69604994 D DE 69604994 T US 5882909 A US 5885810 A US 5807725 A	11-03-1997 09-12-1999 27-04-2000 16-03-1999 23-03-1999 15-09-1998
EP 0674000 A	27-09-1995	JP 7308193 A US 5773273 A	28-11-1995 30-06-1998
WO 0001650 A	13-01-2000	AU 4863099 A EP 1095002 A	24-01-2000 02-05-2001
EP 1063297 A	27-12-2000	JP 2001000192 A	09-01-2001
WO 9727285 A	31-07-1997	US 6087563 A AU 1845397 A BR 9707200 A EP 0877793 A JP 11510708 T	11-07-2000 20-08-1997 28-12-1999 18-11-1998 21-09-1999
WO 9904622 A	04-02-1999	AU 8506198 A EP 1009812 A	16-02-1999 21-06-2000
WO 9906580 A	11-02-1999	AU 8598998 A EP 1002116 A	22-02-1999 24-05-2000

Form PCT/ISA/210 (patent family annex) (July 1992)

INTERNATIONAL SEARCH REPORT
Information on patent family members

Interr. Application No
PCT/US 00/10368

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9906580 A		ZA 9806872 A	02-02-1999
WO 0022150 A	20-04-2000	AU 6290099 A	01-05-2000
WO 0121650 A	29-03-2001	NONE	

Form PCT/ISA/210 (patent family annex) (July 1992)